DOCUMENT RESUME

ED 148 862 TM 006 752

AUTHOR	Braswell, James; Petersen, Nancy
TITLE	An Investigation of Item Obsolescence in the Scholastic Aptitude Test.
INSTITUTION	College Entrance Examination Board, New York, N.Y.
SPONS AGENCY	Educational Testing Service, Princeton, N.J.
PUB DATE	Jan 77
NOTE	79p.; For related documents, see TM 006 750-776 and TM 006 805; For complete report, see TM 006 516
AVAILABLE FROM	College Board Publication Orders, Box 2815, Princeton, New Jersey 08541 (Item No. 251706X, $5.00)
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	College Bound Students; *College Entrance Examinations; *Complexity Level; Educational Problems; Educational Trends; Equated Scores; Item Analysis; Mathematics; *Obsolescence; *Performance Factors; Scores; Secondary Education; Tables (Data); Testing Problems; *Test Items; Time; Trend Analysis; Verbal Ability
IDENTIFIERS	*Scholastic Aptitude Test; *Test Score Decline

ABSTRACT

The decline in Scholastic Aptitude Test (SAT) scores is a direct result of lower average performance on equating questions that are reused over a period of several years. Therefore, it is important to examine the content of representative questions, used over time, for possible obsolescence. Two panels, one concentrating on the verbal section of the SAT and the other on the mathematical, were appointed to review questions that appeared in earlier editions of the test and then in more recent editions. Raters were asked to indicate how the difficulty of each question might be expected to change between the two administration dates. The raters' predictions were then compared with available item analysis data. For the most part, the changes predicted by raters were not substantiated by the statistical analysis. While the relative difficulty of some questions changed between administration dates, it was not possible, except for a few mathematical questions, to attribute these changes to curricular change or to broader social factors. These mathematical questions were predicted by the raters to be easier at the more recent administration, and their prediction was supported by the statistical analysis. The rating form and the resulting data are appended. (Author/EVH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED148862

An Investigation of Item Obsolescence in the Scholastic Aptitude Test

James Braswell
Nancy Petersen
Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
College Entrance Examination Board
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM006 752

Prepared for the Advisory Panel on
the Scholastic Aptitude Test Score Decline,
jointly sponsored by the
College Board and Educational Testing Service

January 1977

This appendix is one of 27 that have been published individually to supplement On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline. The report itself is available at $4 from College Board Publication Orders, Box 2815, Princeton, New Jersey 08540. Additional copies of this appendix, at $5 per copy, as well as a price list for all the appendixes can be obtained from the same address. Payment must accompany any order not submitted on an institutional purchase order.

Copyright © 1977 by College Entrance Examination Board. All rights reserved.

Printed in the United States of America.

ACKNOWLEDGMENTS

The College Board and Educational Testing Service wish to express their appreciation to the following panel members who reviewed verbal and mathematical test questions.

Verbal Panel

William Dwyer
Staff Writer, Trentonian
Trenton, New Jersey

Paul Glatzer
History and Anthropology Teacher
Walt Whitman High School
Huntington Station, L.I., New York

Hugh Lane
Research Associate
Institute for Studies in Education
Washington, D.C.

Jean Malmstrom
Department of English
Western Michigan University
Kalamazoo, Michigan

Sharilyn McSwan
Language Arts Resource Teacher
Grades 7-12, San Diego Public Schools
San Diego, California

Maria Penny
Physics Teacher
Wilde Lake High School
Columbia, Maryland

Charles Suhor
Supervisor of Secondary English
New Orleans Public Schools
New Orleans, Louisiana

Mathematics Panel

Betty Hinman
Division of Mathematics
University of Houston
Houston, Texas

David Johnson
Mathematics Teacher
Nicolet High School
Milwaukee, Wisconsin

Katherine Layton
Mathematics Teacher
Beverly Hills High School
Beverly Hills, California

Willie May
Mathematics Teacher
Wendell Phillips High School
Chicago, Illinois

Eugene Nichols
Department of Mathematics Education
Florida State University
Tallahassee, Florida

William E. Slesnick
Department of Mathematics
Dartmouth College
Hanover, New Hampshire

Robert Smith
Mathematics Teacher
Lawrence High School
Lawrenceville, New Jersey

ABSTRACT

Two panels, one verbal and the other mathematical, were appointed to review and rate questions that appeared in earlier editions of the Scholastic Aptitude Test (SAT) and then in more recent editions. The elapsed time between test administration dates varied between three and 13 years. Raters were sent copies of the various tests and were asked to use a five point rating scale to indicate how the difficulty of each

question might be expected to change between the two administration dates. The raters' predictions were then compared with available item analysis data. Of particular interest were those questions for which the item analysis data showed a substantial change in performance and the raters predicted and could explain the change.

The general trend was for verbal raters to rate questions as being more difficult for recent groups while the trend for mathematical raters was to rate the questions as being somewhat easier for recent groups. Inter-rater correlations were generally low. Rater means were computed for each type of verbal and mathematical question by test. No significant differences were found in the verbal means, but on two of the mathematical tests there were significant differences in rater means due primarily to a higher rating assigned to questions classified as "miscellaneous." A similar analysis was done using the actual item analysis data. On none of the mathematical tests and on only one of the verbal tests was there a significant difference in actual student performance by type of question. The more recent group's performance on the analogies was relatively better than the earlier group's performance.

Because the mathematical questions are more closely related to specific school-based learning than are the verbal questions, one might predict that the mathematical raters' judgments would correlate higher with actual performance based on item analysis data than would those of the verbal raters. Such was the case. The verbal correlations were uniformly low for all raters and for both tests (the greatest was 0.16). For two of the four mathematical tests in which the time spread between administrations was relatively long, several of the mathematical raters' judgments correlated reasonably well with item analysis data (the greatest was 0.66). This finding suggests that curricular changes in mathematics over the past 15 years have had a differential impact on questions administered in the 1960s and then re-administered in the 1970s.

The report also contains 20 specific questions for each of verbal and mathematical that had extreme (either high or low) rater means or were identified as being outliers based on the item analysis data. No verbal questions identified at the extremes in terms of rater means were also identified as extreme outliers in the statistical analyses. Three of the 10 mathematical questions identified as extreme in terms of rater means were also identified as outliers in the statistical analyses.

INTRODUCTION

For the past several years both verbal and mathematical scores on the Scholastic Aptitude Test (SAT) have been declining. The decline is a direct result of lower performance of more recent groups of students on the various equating tests. At each administration of a new edition of the SAT, equating questions that appeared in a previous edition of the test one to six years earlier are used to place raw scores on the new operational test on the 200-800 score scale. These questions are usually placed in a separately timed nonoperational section of the

test. Separate equating is done for the verbal and mathematical parts. The verbal equating tests contain 40 questions that are representative in content and difficulty of the operational verbal test. The mathematical equating tests contain 25 questions that are likewise representative of the operational mathematical test.

Because score decline is a direct result of lower average performance on equating questions that are reused over a period of several years, it is important to examine the content of representative questions used over time for possible obsolescence. Are there factors, curricular or otherwise, that would make certain questions less relevant today than they were in the past? Are there some questions that might be more relevant today?

The primary purpose of this study was to identify and examine those verbal and mathematical questions that changed in difficulty between administration dates. While it has been noted that the mean scores on equating tests have been gradually declining, it has also been observed that performance on some questions remained relatively stable. In contrast to the general trend, a few questions have actually become easier for more recent groups of students.

DESIGN OF THE STUDY

Two independent panels of raters, one for verbal and one for mathematical, were appointed. Raters were sent copies of the various tests and were asked to use a five point rating scale to indicate how the difficulty of each question might be expected to change over the given time period. In addition, the rating form provided space for the rater to comment on the probable cause of the predicted shift in difficulty. The raters' predictions were then compared with the available item analysis data for the two administration dates of each set of questions. Of particular interest were those questions in which the item analysis data showed a substantial change in performance and the raters predicted and could explain the change.

TEST FORMS

Two complete 90 question verbal operational tests and four complete 25 question mathematical equating tests were used in the study. The two verbal tests were those used in the Modu-Stern SAT scale stability study. The earlier and later administration dates are shown in the chart at the top of the next page.

Test	Number of questions	Administration dates	
		Earlier	Later
Verbal 1	90	May 1963	May 1976
Verbal 2	90	March 1973	March 1976
Math 1	25	May 1962	January 1971
Math 2	25	January 1965	December 1973
Math 3	25	April 1971	November 1975
Math 4	25	March 1972	April 1976

TEST REVIEW PANELS

The SAT-verbal is a general test of developed verbal ability. In order to insure that the test discriminates fairly among students, no matter what their particular field of interest, specifications are set so that the content of each test includes material from a variety of fields. Discrete questions, questions that are not related to passages, are taken from the fields of aesthetics and philosophy, human relationships, science, and the world of practical affairs. Each edition of the SAT-verbal has a narrative passage, a humanities passage, a science passage, an argumentative passage, and a social studies passage. It follows, therefore, that the verbal questions should be reviewed by people with different interests in order to insure that comments are broad and far-reaching.

The verbal review panel consisted of seven members--four generalists and three specialists. Those classified as generalists were a high school physics teacher, a high school social studies teacher, a newspaper writer, and an educational researcher. Those classified as specialists were a member of a university English department, a language arts curriculum coordinator, and a curriculum specialist in a large city school system.

Because the mathematical portion of the SAT assumes certain elementary content prerequistes in arithmetic, algebra, and geometry, it was decided that the seven member mathematical panel would consist of four high school mathematics teachers who would be aware of what is and has been taught and emphasized and three specialists at the university level who have been aware of major curricular trends in mathematics over the past 10-15 years. Two of the specialists are authors of textbooks at the high school level. All three specialists are involved in either teaching mathematics, training mathematics teachers, or conducting related research. Two of the panel members, one high school teacher and one college teacher, are currently serving on a panel appointed by the Mathematical Association of America to study the decline in the mathematical preparation of students entering college.

RATING INSTRUCTIONS

Separate detailed instructions for reviewing the test questions and test review forms were prepared for the SAT-verbal and the SAT-mathematical tests. Copies of these instructions are included in Appendix A. Each panel member was asked to consider each question in the various test booklets and to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates were given on the cover of each test booklet. In making each judgment, the raters were asked to assume that the earlier and the later student groups were equal in ability and that the change they predicted in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it was based.

The following codes were used by each panel member to indicate his or her judgment about the change in difficulty for each question:

++ Denotes questions that should be considerably easier at the more recent administration date than at the earlier administration date.

+ Denotes questions that should be somewhat easier at the more recent administration date than at the earlier administration date.

0 Denotes questions that should be approximately the same difficulty.

- Denotes questions that should be somewhat more difficult at the more recent administration date.

-- Denotes questions that should be considerably more difficult at the more recent administration date.

A review form for coding responses was provided for each test. Space was also provided on the form for the reviewers to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. This "Comments" section allowed the reviewers to indicate specific reasons (curricular or societal) for a predicted change in the difficulty level of a question. Reviewers were particularly encouraged to comment on questions coded either ++ or --.

THE DATA

There are two types of data for each test question: (1) the raters' judgments regarding the change in difficulty of the question over time, and (2) a delta value, a measure of the question's difficulty based on the fraction of candidates responding correctly to the question at each administration of the test. We will first discuss the judgments made by the raters and then compare their judgments with the observed changes in question difficulty.

In the analysis to follow, the codes ++, +, 0, -, and --, used by each panel member to indicate his or her judgment about the change in difficulty for a question, have been recoded 5, 4, 3, 2, and 1,

respectively, to enable numerical analysis of the judgments. The raters' judgments for each question within each test are recorded in Tables B1-B2 (see Appendix B). With the exceptions of verbal raters one and three, each rater tended to code most questions 3 indicating that they thought most questions should have been of approximately the same difficulty at each administration. The means and standard deviations of the ratings for each panel member by test and across all tests (two for verbal and four for mathematical) are given in Table B3 (see Appendix B). The average rating for the verbal questions was 2.78 indicating that, overall, the verbal raters felt that candidates should have found the questions slightly more difficult at the more recent administrations. The mean of 2.17 for verbal rater three was considerably lower than the mean for any other verbal rater. The large standard deviation of 1.06 for verbal rater one indicates that this rater tended to use the full range of codes in making judgments about questions.

The average rating of 3.11 for the mathematical questions indicates that, overall, the mathematical raters felt that candidates should have found the mathematical questions slightly easier on the more recent administrations. Mathematics Test 2 contained more questions than the other tests that raters predicted would be relatively easier at the more recent administration (1965-1973). The rater means for this test were generally higher than the means for the other three mathematics tests. The reasons given by raters in the space provided on the rating form were uniformly related to curricular changes--changes that would in most cases benefit students who took the test in 1973. There were some individual questions with low rater means and reasons given for the predicted decline in performance also related to changes in curricular emphases. As can be seen later when individual mathematics questions are examined, there was some consistency of agreement between mean rater judgments and the actual statistical results derived from item analysis data.

The inter-rater correlations by test and across all tests are given in Table B4 (see Appendix B). These correlations excluded items that were extremely easy or extremely hard because such questions are subject to greater chance variation. This is especially true for the more difficult questions because of random guessing. The excluded questions were those with a ** in Tables B5-B10. The inter-rater correlations tend to be small for verbal questions. This is partly due to the lack of variability in the ratings assigned to the various questions. Each panel member correlates positively, indicating agreement, with the mean panel rating. However, there are a few negative correlations between panel members. Some raters tended to code most of the questions on a test 3, indicating no change in difficulty. Consequently, when these raters coded a question other than 3, they disagreed a couple of times with other panel members. These few disagreements tended to cause the small negative correlations. The average inter-rater verbal correlation was 0.12.

The mathematical inter-rater correlations are generally low to moderate. There was considerably more agreement among the mathematical raters than the verbal raters. This is probably due in part to the fact that all seven members of the mathematical panel were in that discipline. The average inter-rater mathematical correlation was 0.25.

The deltas, distance from the major axis, and the mean and standard deviation of panel ratings for each question are given in Tables B5-B10 (see Appendix B). The information in these tables is explained and discussed below.

The difficulty level of the test questions at Educational Testing Service is reported on a "delta" scale. The more difficult the question the greater the delta value. The relationship between delta and the traditional percent correct can be found in Figure B1 (see Appendix B). Except for Verbal Test 2 and Mathematics Test 4, the mean delta for the more recent administration of the tests was larger than that for the earlier administrations, indicating that the more recent groups found the test more difficult. Given that the tests were, on the average, more difficult for the more recent groups, one goal of the study was to identify those questions that were outliers and then to examine the raters' judgments about these questions.

In order to identify outliers, the deltas were plotted so that each point on the graph represents the pair of deltas observed at the two administrations of the question. The plot of these points normally forms an elongated ellipse extending from the lower left corner to the upper right corner. If the candidate groups taking the test at each administration are from similar populations, the ellipse is long and narrow indicating a high correlation. When the groups differ in mean ability level, the ellipse will be displaced above or below the 45° diagonal line (equal-difficulty line) depending on which group is the more able. The greater the discrepancy between the two groups in the way they respond to individual questions, the farther the corresponding points will lie from the major axis line of the ellipse. The perpendicular or shortest distance of the point from the major axis line is a measure of this discrepancy.

To illustrate, the delta plot for Verbal Test 1 is given in Figure B2 (see Appendix B). The horizontal axis represents the more recent administration, and the vertical axis represents the earlier administration. The major axis of the ellipse lies parallel and below the equal-difficulty line (the broken 45° diagonal line) indicating that the more recent group was less able than the earlier group. Points which fall on the major axis indicate questions which probably represent the same task for both groups. Points above the major axis indicate questions which were relatively more difficult for the earlier group and have negative perpendicular distances associated with them, whereas those below the line indicate questions which were relatively more difficult for the more recent group and have positive perpendicular distances associated with them.

Not all questions were used in the construction of the delta plots. The usefulness of a delta plot depends on the soundness of the assumption that the major axis of the ellipse adjusts for differences in ability level between the two groups and that its location is unaffected by other confounding factors such as item obsolescence or error generated by guessing. In an attempt to remove any such biases, all questions for which the mean rating was greater than 3.3 or less than 2.7 were deleted from construction of the delta plot, as such ratings might be indicative of questions that have changed in meaning. Also, all questions for which the delta values in at least one administration were greater than 15.7 or less than 6.4 were deleted. A delta of less than 6.4 corresponds to 95 percent or more of the candidates responding

correctly to the question whereas a delta greater than 15.7 corresponds to 25 percent or less of the candidates responding correctly, that is, the guessing range. This last group of questions were deleted from all remaining analyses of the data.

ANALYSIS OF THE DATA

Each of the verbal questions is categorized as either sentence completion, antonym, analogy, or reading passage. Each of the mathematical questions is categorized as either arithmetic, algebra, geometry, or miscellaneous. It is of interest to determine whether (1) the raters judged particular types of questions as becoming easier or more difficult, (2) the perpendicular distances indicate that particular types of questions have become easier or more difficult, and (3) whether the two indices agree. It is also of interest to determine whether the raters or the perpendicular distances indicate more change in question difficulty as the time between administrations of the questions increases.

The one-way analysis of variance technique was used to study the above questions for each test. The one-way analysis of variance technique is a method for comparing means when there are two or more categories of interest. The hypothesis being considered is that the means for each category are identical. The test statistic used is called the F ratio. The F ratio is basically the ratio of the variance between the means for each category to the variance between observations within each category. If the hypothesis is false, that is, if the means for the categories really do differ, then the differences among the means will be large in relation to the differences among observations in each category and the resulting F ratio will be large. The larger the F ratio the smaller the probability that the means for each category differ only by chance.

Table 1 summarizes the one-way analysis of variance results for the panels' average question ratings categorized by question type. Except for Mathematics Tests 2 and 3, none of the F ratios are large enough to say that the differences in the means for the various types of questions are due to something other than chance, that is, there is not enough evidence to say that raters tend to view any particular type of question as having changed in difficulty over time. However, for Mathematics, there is a trend. The average rating for the "miscellaneous" questions is always the largest, and being greater than 3, this indicates that the raters thought these questions would be easier at the more recent administration.

The one-way analysis of variance results for the perpendicular distances categorized by question type are summarized in Table 2. Except for Verbal Test 2, the F ratios are nonsignificant. This time there is no consistent trend in the means for the mathematical questions. However, there is a trend for the verbal questions. The analogy type question appears to have been easier at the more recent administration (negative mean perpendicular distance). Thus, neither the analysis of the rater data nor the analysis of the distance data consistently identified types of questions such as analogy or algebra as having changed in meaning between test administrations.

Table 1. One-Way Analysis of Variance
Average Rating Categorized by Question Type

Verbal

Test	Summary statistics	Question type: Sentence completion	Antonym	Analogy	Reading passage	F ratio
1	n	16	12	17	29	
	mean	2.77	2.63	2.78	2.87	1.40
	sd	.40	.28	.24	.36	
2	n	12	13	14	33	
	mean	2.95	2.91	2.79	2.80	2.16
	sd	.22	.12	.25	.21	

Mathematics

Test	Summary statistics	Question type: Arithmetic	Algebra	Geometry	Misc.	F ratio
1	n	8	2	6	0	
	mean	3.02	2.86	3.05		.34
	sd	.31	.29	.16		
2	n	4	9	5	4	
	mean	3.07	3.30	3.31	4.00	4.08*
	sd	.50	.33	.21	.43	
3	n	5	7	3	2	
	mean	3.03	3.14	3.05	3.36	3.70*
	sd	.06	.11	.18	.07	
4	n	5	6	6	3	
	mean	3.08	3.07	3.07	3.19	.84
	sd	.11	.07	.14	.07	

*Significant at 5 percent level.

Table 2. One-Way Analysis of Variance
Distance from Major Axis Categorized by Question Type

Verbal

Test	Summary statistics	Question type: Sentence completion	Antonym	Analogy	Reading Passage	F ratio
1	n	16	12	17	29	
	mean	.01	.03	-.28	.03	1.56
	sd	.37	.57	.55	.47	
2	n	12	13	14	33	
	mean	.06	.04	.17	.04	2.75*
	sd	.24	.27	.34	.16	

*Significant at 5 percent level.

Mathematics

Test	Summary statistics	Question type: Arithmetic	Algebra	Geometry	Misc.	F ratio
1	n	8	2	6	0	
	mean	-.03	-.28	.09		.58
	sd	.31	.38	.46		
2	n	4	9	5	4	
	mean	.35	-.19	-.01	-.20	1.16
	sd	.42	.57	.23	.42	
3	n	5	7	3	2	
	mean	.09	-.07	.09	-.15	.56
	sd	.33	.23	.19	.14	
4	n	5	6	6	3	
	mean	.07	-.16	-.07	.33	2.24
	sd	.16	.20	.37	.09	

The time between the two administrations for the various tests differs substantially. Thirteen years elapsed between the two administrations of Verbal Test 1 whereas only three years elapsed between the two administrations of Verbal Test 2. Similarly, nine and eight years elapsed between the two administrations of Mathematics Tests 1 and 2, respectively, whereas only four years elapsed between the two administrations of Mathematics Tests 3 and 4. Perhaps, by analyzing the data by time between administrations, one can determine whether the raters thought there should be more change in the difficulty level of a question as the number of years between administration increases.

Table 3 summarizes the one-way analysis of variance results for the panels' average ratings categorized by tests. Mathematics Tests 1 and 2 and Mathematics Tests 3 and 4 were combined for each analysis as the time span between the two administrations for each set of tests was approximately the same. Neither of the F ratios is large enough to suggest that the difference in the mean ratings are due to something other than chance.

Since the analyses using the ratings and the analyses using the distances do not suggest the same sort of trends, it is of interest to see just how much agreement there is between the two measures. Table 4 summarizes the one-way analysis of variance results for perpendicular distances categorized by average rating. If the two types of data agree, then those questions with a rater mean greater than 3 should have a negative mean perpendicular distance, those questions with a rater mean of 3 should have a mean perpendicular distance of approximately zero, and those questions with a rater mean less than 3 should have a positive mean perpendicular distance. The F ratios were nonsignificant, that is, there is not enough evidence to say that the

Table 3. One-Way Analysis of Variance
Average Rating Categorized by Test

Summary statistics	Verbal		F
	Test 1	Test 2	ratio
n	74	72	
Mean	2.79	2.84	1.40
sd	.34	.22	

Summary statistics	Mathematics		F
	Test 1 and 2	Test 3 and 4	ratio
n	38	37	
Mean	3.23	3.10	2.66
sd	.44	.13	

Table 4. One-way Analysis of Variance
Distance from Major Axis Categorized by Average Rating

Verbal

Test	Summary statistics	Average rating			F ratio
		Greater than 3	Equal to 3	Less than 3	
1	n	13	13	48	
	mean	.22	-.17	-.08	2.47
	sd	.60	.60	.41	
2	n	11	17	44	
	mean	-.01	.02	-.00	.05
	sd	.17	.25	.27	

Mathematics

Test	Summary statistics	Average rating			F ratio
		Greater than 3	Equal to 3	Less than 3	
1	n	7	3	6	
	mean	-.32	.14	.26	1.48
	sd	.34	.15	.27	
2	n	17	2	3	
	mean	-.17	.67	.13	3.18
	sd	.43	.16	.56	
3	n	9	7	1	
	mean	-.06	.10	-.16	.84
	sd	.24	.28	.00	
4	n	13	5	2	
	mean	.09	-.16	-.23	2.04
	sd	.29	.10	.40	

means differ other than by chance. Also, the hypothesized trend in the means described above is not even present in each data set.

For an item to have a rater mean greater than 3, for example, it is not necessary for all raters to rate the question as either 3, 4, or 5. It is possible, then, that lack of rater consistency is clouding the results. To see if this is the case, an analysis similar to that summarized in Table 4 was done, using only those items for which the raters' judgments were consistently 3 or more, 3 or less, or equal to 3. However, again the F ratios were nonsignificant.

The correlations of rater judgments on individual questions with the distance of that question from the major axis are given in Table B4 for each test. This data is derived by correlating the judgments of each rater in Tables B1 and B2 with the corresponding distance values in Tables B5-B10. Recall that a question rated 5 is judged to be considerably easier for the more recent group while a question rated 1 is judged to be considerably more difficult. Also, recall that outliers that have negative distances associated with them are relatively more difficult for the earlier group while those with positive distances associated with them are relatively more difficult for the recent group. Therefore, a rater judgment of 5 would be consistent with an outlier distance value of -1.1 and a rater judgment of 2 would be consistent with a distance value of +0.4. Thus, a large negative correlation indicates agreement.

The verbal correlations are generally quite small and sometimes in the opposite direction from what one would expect. In fact, the correlation between the average rating and the distance from the major axis on Verbal Test 1 is 0.13 and on Verbal Test 2 is 0.01 which indicate slight disagreement between means and distances.

Some of the mathematical correlations are surprisingly high. For example, on Mathematics Test 1, the average rating and raters one and two had correlations of -0.60, -0.57, and -0.56, respectively, with the distances and on Mathematics Test 2, the average rating and raters one and five had correlations of -0.61, -0.66, and -0.44, respectively, with the distances, indicating a reasonable amount of agreement between raters' judgments and actual distance results. The correlations for Mathematics Tests 3 and 4, on the other hand, are uniformly small.

Given the rather large correlations between average rating and distance from major axis on Mathematics Tests 1 and 2, it seems somewhat surprising that the result of the analysis in Table 4 is nonsignificant. Figures B3 and B4 are scatterplots of the average rating versus the distance from the major axis for Mathematics Tests 1 and 2, respectively. In Figure B3, questions 4, 7, 13, and 14 appear to be outliers. If these four questions are deleted, the correlation decreases from -0.60 to -0.13. In Figure B4, questions 2 and 11 appear to be outliers. If these two questions are deleted, the correlation decreases from -0.44 to -0.11. Thus, the rather large correlations between average rating and distance from major axis on Mathematics Test 1 and 2 seem to be largely due to the presence of a few questions which were outliers having consistent extreme average ratings and extreme distance values.

A LOOK AT PARTICULAR QUESTIONS

This section of the report will examine those specific questions that had either high mean rankings, low mean rankings, or were identified as being outliers on the delta plots. The raters made many extremely interesting comments on individual questions. Although most of the comments were not confirmed by the delta plots and F-tests, many of the comments will be useful in future test assembly.

Tables C1-C2 in Appendix C present those five verbal questions that had the highest rater means and those five questions that had the lowest rater means. Questions with a high rater mean represent questions that the raters predicted would become relatively easier for the more recent groups while those with low rater means would become harder. Included for each question are the rater mean, the perpendicular distance, and the specific comments made by raters. For those questions that relate to a reading passage, the entire passage is shown.

Tables C3-C4 present the same type of information for the five verbal questions at the greatest negative distances from the major axis and the five questions at the greatest positive distances. The negative distances favor the more recent groups while the positive distances favor the earlier groups. No question identified in Table C1 is also identified in Table C3. This means that none of the top five questions judged by the rater means as most likely favoring the more recent groups were in the top five questions identified in the delta plot analysis. Likewise, there were no common questions identified in Tables C2 and C4. Thus, none of the top five verbal questions identified by the rater means as favoring the earlier groups were identified in the corresponding delta plot analysis. There are, however, some trends that run in the same direction. On Test 1 the passage for questions 86-90 was computer-related. Several of the raters felt that today's students have greater knowledge of computers and related vocabulary. Questions 87 and 90 had relatively high rater means and each was at a negative distance of -0.49 from the major axis, indicating these questions were in fact relatively easier for the more recent group.

It should be noted that questions at the extremes of the difficulty scale were considered in this portion of the analysis and that chance factors alone could be responsible for observed delta changes on difficult questions. Most of the extremely difficult vocabulary questions that raters generally ranked as more difficult for the recent group actually turned out to be relatively less difficult. A primary reason for this could have been the random noise generated by guessing.

Tables C5-C6 in Appendix C show those five mathematical questions that had the highest rater means and those five questions with the lowest rater means. Tables C7-C8 show those five mathematical questions that had the greatest (negative) perpendicular distances from the major axis and the five questions that had the greatest (positive) perpendicular distances. Of the five questions with the highest rater means (indicating questions that should be easier for more recent groups), two of these questions were also identified as extremes in the delta plots. These were questions 5 and 11 on Test 2. Of the five mathematical questions with the least rater means (indicating questions that should be harder for more recent groups), only one of these questions was also identified as an extreme by the delta plots. This was question 2 on Test 2.

There were two areas where the mathematical raters' opinions were supported by the delta plots. The raters stated for questions 5 and 11 on Test 2 that greater recent emphasis on real number properties, properties of zero, and abstract operations, should have made these questions easier for more recent groups. These questions were also identified in the delta plots as questions that were relatively easier

for the more recent group. Question 2 on Test 2 which involves percent, was identified in the top five by both raters and delta plots as being more difficult for the recent group. The rater comments (see Table C5, Appendix C) do not suggest why the question would be more difficult, but comments by raters on other percent questions suggest that this topic is receiving less emphasis on the curriculum.

IMPLICATIONS OF THE STUDY

In order to relate raw scores on each new edition of the SAT to the established 200-800 score scale, separately timed verbal and mathematical "equating" tests are readministered in a nonoperational section of the SAT. Through the equating process, raw scores on the new test are adjusted for the estimated difference between the ability of the group that took the new test and the ability of a prior reference group. A typical equating test is used three to five times over a five year period and then retired. The equating process assumes that questions readministered represent the same psychological task at each administration. This study was conducted to determine whether there are intervening curricular or societal factors that might produce equating error.

Representative verbal and mathematical questions were rated by two independent panels and no evidence was found to suggest that identifiable curricular and societal factors are responsible for short term (three to five years) observed changes in item performance. The same was true for verbal items, whose administration dates were 13 years apart. However, for the mathematical questions, there is evidence to suggest that identifiable curricular factors may account for the differential performance on a few of the questions. In view of this evidence, the current practice of retiring equating tests after about five years seems appropriate, especially for the mathematical equating tests.

SUMMARY

Score decline is a direct result of lower average performance on equating questions that are reused over a period of time. It is possible that the lower performance might be explained by curricular or societal changes that would make certain questions less relevant today than they were in the past. The primary purpose of this study, then, was to identify those verbal and mathematical questions that have changed substantially in difficulty between administrations.

Two 90 question verbal operational tests and four 25 question mathematical equating tests were used in the study. Questions were identified as outliers using (1) judgments made by raters and (2) perpendicular distances calculated from delta-plots.

The ratings and the perpendicular distances were both analyzed by type of question. Analysis of the ratings suggested that the raters

thought that mathematical "miscellaneous" questions would have been easier at the more recent administrations. Analysis of the perpendicular distances indicated that verbal analogy questions tended to be easier at the more recent administrations. The conclusions from the two analyses by type of question do not tend to support each other. The ratings were also analyzed by elapsed time between test administrations. The analysis failed to suggest that differences in rater means for the two time periods were due to something other than change.

Analysis of the perpendicular distances classified by average rating showed that raters were, on the whole, unable to predict the observed change in difficulty level of a question between administrations. This result seemed somewhat incongruous with the high correlations between average rating and distance from major axis for Mathematics Tests 1 and 2. However, examination of the scatter plots showed that the high correlations of average rating with distance from major axis for Mathematics Tests 1 and 2 were largely due to a few questions that were outliers having consistent extreme average ratings and extreme distance values.

Also examined were each of the five verbal and mathematical questions with (1) the highest mean rating, (2) the lowest mean rating, (3) the greatest negative perpendicular distance, and (4) the greatest positive perpendicular distance. None of the verbal questions and only three of the mathematical questions were identified as outliers in the same direction by both sets of data.

This study was conducted to see whether raters might be able to predict and explain changes in the difficulty level of questions over time. For the most part, the changes predicted by raters were not substantiated by the statistical analysis. Therefore, while the relative difficulty of some questions changed between administration dates, it was not possible, except for a few mathematical questions, to attribute these changes to curricular or societal factors.

APPENDIX A:

INSTRUCTIONS TO PANELS AND RATING FORM

INSTRUCTIONS FOR REVIEWING VERBAL TEST QUESTIONS

Educational Testing Service is conducting a study on behalf of the College Board to determine whether or not certain questions that have been administered in two previous editions of the Scholastic Aptitude Test (SAT) represent different tasks when given at different points in time. It is conceivable that, due to curricular or social changes, some questions became more (or less) difficult between the first and the second test administrations.

Consider, for example, how two reading selections, one about the blitzkriegs of the Second World War and one about the problems of weightlessness in space, would have reversed as far as familiarity of content is concerned in the last thirty years. It is also conceivable that some reading selections and questions based on them might be as appropriate today as they were in the past, but changing attitudes or new knowledge may change the attractiveness of incorrect multiple-choice alternatives and thus affect the difficulty of the questions.

With the passing of time, certain words may change in meaning or become obsolete. Other words may become more familiar due to more frequent use. Consider this analogy item:

CISTERN:WATER:: (A) shower:cloud (B) official:power
(C) science:matter (D) museum:antiques (E) vault:valuables

It seems likely that the difficulty of this question would be greater today than in the past because fewer of today's students would know the word "cistern."

For each question in the enclosed test booklets, you are to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates are given on the cover of the test booklet. You are asked to consider the entire question, including the stem, the alternatives, and the material on which the question is based. In making each judgment, you should assume that the earlier and later student groups were equal in ability and that the change you note in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it is based.

Please use the following codes to indicate your judgment about the change in difficulty for each question:

+ + Denotes questions that you think should be <u>considerably easier</u> (much higher percentage of students respond correctly) at the more recent administration date than at the earlier administration date (dates given on test copy).

+ Denotes questions that you think should be <u>somewhat easier</u> at the more recent administration date than at the earlier administration date.

0 Denotes questions that you think should be <u>approximately the same difficulty</u> at the more recent administration date as at the earlier administration date.

- Denotes questions that you think should be somewhat more difficult at the more recent administration date than at the earlier administration date.

- - Denotes questions that you think should be considerably more difficult (much lower percentage of students respond correctly) at the more recent administration date than at the earlier administration date.

Use the Review Form to record your responses. A set of 4 Review Forms are stapled together to correspond to each test booklet. An extra set is included. Be sure to write the Test Number (given on the cover of the test booklet) in the space provided on the Review Form. Columns are included on the Review Form for you to check the appropriate code. Columns are also provided for you to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. Leave these columns blank if you do not know whether curricular changes affected the question you are rating. The "Comments" section allows you to indicate the specific reasons (curricular or societal) why you think the difficulty level of a question changed. It is particularly important that a brief remark be made for questions coded either + + or - - .

Before you begin your review, you should scan the enclosed pamphlet "About the SAT." It will probably be easiest for you to make the judgments asked for if you take the tests yourself and make your judgments as you go. Answer keys are provided for your reference.

After you have completed the Review Form,

- Check to make sure you have written in the Test Number, and

- Sign your name in the upper right-hand corner of each Review Form.

Then

- Fill out the social security form (needed by the ETS Payroll Department);

- Return the social security form, test booklets, answer keys, and rating forms in the envelope provided. You may wish to copy your rating form before returning it.

Thank you for your help and cooperation!

INSTRUCTIONS FOR REVIEWING MATHEMATICAL TEST QUESTIONS

Educational Testing Service is conducting a study on behalf of the College Board to determine whether or not certain questions that have been administered in previous editions of the Scholastic Aptitude Test (SAT) represent different tasks when given at different points in time. It is conceivable that, due to curricular or social changes, some questions became more (or less) difficult between the first and the second test administrations.

For example, those questions that are more in the spirit of contemporary mathematics might be expected to be easier for more recent groups of secondary school students. However, those questions that refer to topics or settings that were more familiar to students in the mid-1960's than to recent groups might have become more difficult.

For each question in the enclosed test booklets, you are to judge whether or not students would have found the question easier or harder at the second administration than at the first administration. The administration dates are given on the cover of the test booklet. You are asked to consider the entire question, including the stem, the alternatives, and the material on which the question is based. In making each judgment, you should assume that the earlier and later student groups were equal in ability and that the change you note in difficulty level was due not to any change in the ability of the students, but to a change in the meaning of the question or the familiarity of the material on which it is based.

Please use the following codes to indicate your judgment about the change in difficulty for each question:

++ Denotes questions that you think should be <u>considerably easier</u> (much higher percentage of students respond correctly) at the more recent administration date than at the earlier administration date (dates given on test copy).

\+ Denotes questions that you think should be <u>somewhat easier</u> at the more recent administration date than at the earlier administration date.

0 Denotes questions that you think should be <u>approximately the same difficulty</u> at the more recent administration date as at the earlier administration date.

\- Denotes questions that you think should be <u>somewhat more</u> difficult at the more recent administration date than at the earlier administration date.

-- Denotes questions that you think should be <u>considerably more difficult</u> (much lower percentage of students respond correctly) at the more recent administration date than at the earlier administration date.

Use the Review Form to record your responses. Fill out one Review Form for each test booklet. Be sure to write the Test Number (given on the cover of the test booklet) in the space provided on the Review Form. Columns are included on the Review Form for you to check the appropriate code. Columns are also provided for you to indicate the extent to which curricular changes may have contributed to changes in difficulty during the time between the two administration dates. Leave these columns blank if you do not know whether curricular changes affected the question you are rating. The "Comments" section allows you to indicate the specific reasons (curricular or societal) why you think the difficulty level of a question changed. It is particularly important that a brief remark be made for questions coded either + + or - - .

Before you begin your review, you should scan the enclosed pamphlet "About the SAT." It will probably be easiest for you to make the judgments asked for if you take the tests yourself and make your judgments as you go. Answer keys are provided for your reference.

After you have completed the Review Form,

- Check to make sure you have written in the Test Number, and
- Sign your name in the upper right-hand corner of each Review Form.

Then

- Fill out the social security form (needed by the ETS Payroll Department);
- Return the social security form, test booklets, answer keys, and rating forms in the envelope provided. You may wish to copy your rating form before returning it.

Thank you for your help and cooperation!

REVIEW FORM

Rater _____

Test No. _____

Question Number	Difficulty for Recent Group (Check One)					Effect of Curricular Changes on the Judged Change in Difficulty Over Time (Check One)			COMMENTS (This space can be used to explain your coded response. It is particularly important that a brief remark be made for questions coded either + + or - - .)
	Easier		=	Harder		Considerable Effect	Some Effect	Little or No Effect	
	++	+	0	-	--				
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									

APPENDIX B:

DATA, TABLES, AND FIGURES

Table B1. Ratings on Verbal Questions

Question	Verbal test 1 rater							Verbal test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	5	3	3	3	3	3	3	3	3	3	3	3	3	2
2	2	2	1	3	2	3	2	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	4	3	3	3	3	3	2
4	3	3	3	2	3	3	3	2	3	1	3	3	3	3
5	1	2	3	3	3	3	1	2	3	2	3	3	3	3
6	2	3	3	3	2	3	1	3	3	3	3	3	3	3
7	1	3	2	3	3	3	2	5	3	2	3	3	3	3
8	2	3	3	3	3	3	1	3	3	2	3	3	3	3
9	5	3	1	3	3	3	3	3	3	2	3	3	3	3
10	4	3	2	4	3	2	1	2	3	1	3	3	3	3
11	3	3	2	3	3	3	1	2	3	3	2	3	3	3
12	3	3	2	3	3	3	3	4	3	1	3	3	3	2
13	2	3	2	3	3	2	3	4	3	2	3	3	3	3
14	2	3	3	2	3	3	3	4	3	3	3	3	3	3
15	1	3	2	3	3	3	3	4	3	1	3	3	3	3
16	1	3	1	2	3	3	1	4	3	3	3	3	3	3
17	4	3	1	3	3	3	3	4	3	3	3	3	3	3
18	1	3	1	3	3	3	1	3	3	2	3	3	3	3
19	2	3	1	3	3	3	1	4	3	2	2	3	2	3
20	1	3	1	3	3	3	1	3	3	1	3	3	3	3
21	3	3	3	3	3	3	3	3	3	3	3	3	3	3
22	3	3	3	3	3	3	2	2	3	2	3	3	3	3
23	2	3	3	3	3	3	3	4	3	2	2	3	3	3
24	3	3	3	3	3	3	3	4	3	2	3	3	3	3
25	2	3	2	2	3	3	3	2	3	1	3	3	3	2
26	4	2	2	3	3	3	3	4	3	3	3	3	3	3
27	2	4	3	2	3	3	4	4	3	2	3	3	3	3
28	4	3	1	3	3	3	2	3	3	2	3	3	3	3

(Continued)

Question	Verbal test 1 rater							Verbal test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
29	3	3	3	2	3	3	1	1	3	1	3	3	3	3
30	5	3	3	3	3	3	1	2	3	2	3	3	3	3
31	2	3	3	3	3	3	2	4	3	1	3	3	3	3
32	3	3	2	2	3	3	2	3	3	1	3	3	3	3
33	3	3	3	3	3	3	2	3	3	2	3	3	3	3
34	1	3	3	3	3	3	2	4	3	2	3	3	3	3
35	2	3	3	3	3	2	2	4	3	2	3	3	3	3
36	3	3	1	3	3	3	3	3	3	1	3	3	3	3
37	2	3	3	3	3	3	3	2	3	1	3	3	3	3
38	2	3	2	3	3	3	3	2	3	1	3	3	3	3
39	3	3	3	3	3	3	3	2	3	1	2	3	3	3
40	3	3	3	3	3	3	3	2	3	1	2	3	3	3
41	4	3	3	3	3	3	3	4	3	3	3	3	3	3
42	5	3	3	3	3	3	3	4	3	3	3	3	3	3
43	4	3	3	3	3	3	3	4	3	3	3	3	3	3
44	4	3	1	3	3	3	3	4	3	3	3	3	3	3
45	2	3	3	3	3	3	3	3	3	3	3	3	3	2
46	2	3	1	2	3	2	2	2	3	2	3	3	3	3
47	4	3	3	3	3	2	3	3	3	3	3	3	3	3
48	4	3	3	3	3	2	3	2	3	1	3	3	3	3
49	4	3	3	3	3	2	2	3	3	3	3	3	3	3
50	5	3	3	3	3	2	3	3	3	2	3	3	3	3
51	3	3	3	3	3	3	3	4	3	3	3	3	3	3
52	4	3	2	3	3	3	3	5	3	3	3	3	4	3
53	5	3	3	3	3	3	3	3	3	3	3	3	3	3
54	1	3	3	3	3	3	2	4	3	2	3	3	3	1
55	5	3	3	3	3	3	4	2	3	2	3	3	3	2
56	2	3	3	3	3	3	3	4	3	2	3	3	3	3
57	2	3	1	3	3	3	1	2	3	2	3	3	3	3
58	3	3	1	1	3	3	1	2	3	2	3	3	3	3
59	4	3	3	3	3	3	3	3	3	3	3	3	3	3
60	2	3	3	3	3	3	3	3	3	3	3	3	3	3
61	2	3	2	3	3	3	3	2	3	3	3	3	3	3
62	2	3	1	3	3	3	3	2	3	2	3	3	3	3
63	1	3	1	2	3	3	3	3	3	2	3	3	3	3

(Continued)

Table B1 (Continued)

Question	Verbal test 1 rater							Verbal test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
64	2	3	3	3	3	3	2	3	3	2	3	3	3	3
65	4	3	3	3	3	3	3	2	3	1	3	3	3	2
66	1	3	2	3	3	3	2	1	3	2	3	3	3	2
67	3	3	3	3	3	3	2	2	3	3	3	3	3	3
68	3	3	3	3	3	3	3	3	3	2	3	3	3	3
69	4	3	3	3	3	3	3	2	3	2	3	3	2	2
70	2	3	3	3	3	3	3	3	3	1	2	3	3	1
71	2	3	3	3	3	3	2	2	3	2	3	3	3	3
72	4	3	2	1	3	3	1	2	3	3	3	3	3	3
73	4	3	2	1	3	3	1	2	3	1	3	3	3	3
74	3	3	1	2	3	3	1	2	3	1	3	3	3	3
75	3	3	3	3	3	3	3	1	3	1	2	3	3	3
76	2	3	1	2	3	3	1	2	3	1	3	3	3	3
77	2	3	3	2	3	3	1	2	3	1	3	3	3	3
78	2	3	2	2	3	3	1	2	3	2	2	3	3	3
79	2	3	2	2	3	3	1	2	3	2	3	3	3	3
80	2	3	2	1	3	3	1	2	3	1	2	3	3	3
81	4	3	3	3	3	3	2	4	3	2	3	3	3	3
82	3	3	2	3	3	3	2	2	3	2	3	3	3	3
83	4	3	2	3	3	3	2	4	3	3	3	3	3	3
84	4	3	3	2	3	3	2	4	3	2	3	3	3	3
85	4	3	1	2	3	3	2	4	3	2	3	3	3	1
86	4	4	2	3	3	4	4	2	3	2	3	3	3	3
87	5	3	2	5	3	4	4	2	3	2	3	3	3	3
88	2	3	2	4	3	4	4	2	3	1	3	3	3	3
89	2	3	2	5	3	4	4	2	3	2	3	3	3	3
90	4	3	2	4	3	4	4	3	3	2	3	3	3	2
Mean	2.87	2.99	2.32	2.81	2.98	2.97	2.38	2.89	3.00	2.02	2.89	3.00	2.99	2.82
S.D.	1.17	0.24	0.79	0.66	0.15	0.38	0.93	0.94	0.00	0.75	0.31	0.00	0.18	0.46

Table B2. Ratings on Mathematical Questions

Question	Mathematics test 1 rater							Mathematics test 2 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	2	3	3	3	3	3	3	4	3	3	3	3	3	3
2	4	3	3	3	3	3	3	2	2	3	2	2	3	2
3	2	3	4	3	3	3	3	4	4	3	3	3	3	3
4	3	4	3	3	3	3	4	4	4	3	2	3	3	3
5	2	3	3	3	4	3	4	4	3	4	5	4	3	3
6	2	3	3	3	3	3	3	3	4	3	2	3	3	3
7	4	4	3	4	4	3	3	2	2	4	3	3	3	3
8	4	2	3	3	3	3	3	2	3	4	3	3	3	3
9	2	3	3	3	3	3	3	4	3	3	4	4	3	4
10	2	2	3	2	3	3	2	4	5	4	2	3	3	3
11	3	2	4	3	3	3	3	5	5	4	5	4	4	4
12	2	2	3	2	3	3	3	3	4	3	3	3	3	3
13	3	5	3	3	3	3	3	4	3	2	2	3	3	3
14	5	3	3	3	3	3	2	4	4	4	4	3	3	4
15	2	3	3	3	3	3	3	3	3	4	3	3	3	3
16	3	4	3	3	3	3	3	4	4	3	4	3	4	3
17	2	1	3	2	2	3	3	4	5	4	5	3	4	4
18	3	4	3	3	3	3	3	4	5	4	5	3	4	4
19	2	5	4	3	2	3	3	4	4	4	4	3	3	3
20	1	2	4	3	3	3	2	2	4	4	4	3	3	3
21	4	2	3	3	2	3	3	4	4	3	4	3	3	4
22	3	1	3	3	3	3	3	3	4	3	3	3	3	3
23	2	2	4	2	3	3	2	3	4	3	5	3	4	3
24	4	4	3	3	2	3	3	4	3	3	3	3	2	4
25	3	3	3	3	3	3	3	3	3	4	4	3	3	4
Mean	2.76	2.92	3.20	2.88	2.92	3.00	2.92	3.48	3.68	3.44	3.48	3.08	3.16	3.28
S.D.	0.95	1.06	0.40	0.43	0.48	0.00	0.48	0.81	0.84	0.57	1.02	0.39	0.46	0.53

Table B2 (Continued)

Question	Mathematics test 3 rater							Mathematics test 4 rater						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	3	3	3	3	3	3	3	3	3	3	4	3	3	3
2	3	3	3	3	3	3	3	3	3	4	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	3	4	3	3	3	3	2	3	3	3	3	3	3
5	3	3	4	3	3	3	3	3	3	4	3	3	3	3
6	3	2	3	3	3	3	3	4	3	3	3	3	3	3
7	2	4	4	3	3	3	4	3	3	4	4	3	3	3
8	4	4	4	3	2	3	3	3	3	3	3	3	3	3
9	3	3	4	3	3	3	3	3	3	4	3	4	3	3
10	3	3	4	3	3	3	3	3	3	4	3	3	3	3
11	4	3	3	3	3	3	4	3	3	3	3	4	3	3
12	3	3	4	2	4	4	4	3	3	3	3	4	3	3
13	3	3	3	3	3	3	3	3	3	4	2	3	4	3
14	3	3	4	2	3	3	3	2	3	4	3	2	3	3
15	3	3	3	3	3	3	3	3	3	3	3	3	3	3
16	3	3	4	2	3	3	4	3	3	4	3	3	3	3
17	3	3	3	3	3	3	3	3	3	3	3	3	3	3
18	2	3	3	3	3	4	4	3	3	4	3	3	3	3
19	3	3	3	3	3	3	3	3	3	2	3	3	3	3
20	3	3	3	3	3	3	3	3	3	4	3	3	3	3
21	3	2	3	3	3	3	4	3	3	4	3	3	3	3
22	3	3	4	3	3	3	4	2	3	3	3	3	3	2
23	2	3	4	3	2	3	3	2	3	4	3	3	3	3
24	3	2	3	3	2	3	3	2	3	3	3	3	3	3
25	1	2	3	3	3	3	3	3	3	3	3	2	3	3
Mean	2.92	2.92	3.44	2.88	2.92	3.08	3.28	2.84	3.00	3.44	3.04	3.04	3.04	2.96
S.D.	0.63	0.48	0.50	0.33	0.39	0.27	0.45	0.46	0.00	0.57	0.34	0.45	0.20	0.20

Table B3. Rater Means and Standard Deviations

	Verbal					
	Test 1		Test 2		Total	
Rater	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	2.87	1.17	2.89	0.94	2.88	1.06
2	2.99	0.24	3.00	0.00	2.99	0.17
3	2.32	0.79	2.02	0.75	2.17	0.78
4	2.81	0.66	2.89	0.31	2.85	0.52
5	2.98	0.15	3.00	0.00	2.99	0.10
6	2.97	0.38	2.99	0.18	2.98	0.30
7	2.38	0.93	2.82	0.46	2.60	0.76
Ave.	2.76	0.35	2.80	0.23	2.78	0.30

	Mathematics					
	Test 1		Test 2		Test 3	
Rater	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	2.76	0.95	3.48	0.81	2.92	0.63
2	2.92	1.06	3.68	0.84	2.92	0.48
3	3.20	0.40	3.44	0.57	3.44	0.50
4	2.88	0.43	3.48	1.02	2.88	0.33
5	2.92	0.48	3.08	0.39	2.92	0.39
6	3.00	0.00	3.16	0.46	3.08	0.27
7	2.92	0.48	3.28	0.53	3.28	0.45
Ave.	2.94	0.29	3.37	0.45	3.06	0.19

	Mathematics			
	Test 4		Total	
Rater	Mean	S.D.	Mean	S.D.
1	2.84	0.46	3.00	0.79
2	3.00	0.00	3.13	0.78
3	3.44	0.57	3.38	0.52
4	3.04	0.34	3.07	0.65
5	3.04	0.45	2.99	0.44
6	3.04	0.20	3.07	0.29
7	2.96	0.20	3.11	0.47
Ave.	3.05	0.15	3.11	0.33

Table B4. Intercorrelations

Verbal test 1

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0436	0.0549	0.1147	0.1362	-0.0373	0.2413	0.6207	0.0983
2	0.0436	1.0000	0.1409	-0.1193	0.3414	-0.0045	0.1940	0.2216	-0.1600
3	0.0549	0.1409	1.0000	0.1915	0.0899	-0.1854	0.1782	0.4568	0.1993
4	0.1147	-0.1193	0.1915	1.0000	-0.0513	0.3192	0.5965	0.6633	0.0808
5	0.1362	0.3414	0.0899	-0.0513	1.0000	-0.0065	0.1825	0.2457	0.1345
6	-0.0373	-0.0045	-0.1854	0.3192	-0.0065	1.0000	0.2388	0.2524	-0.0910
7	0.2413	0.1940	0.1782	0.5965	0.1825	0.2388	1.0000	0.7848	0.0570
Ave.	0.6207	0.2216	0.4568	0.6633	0.2457	0.2524	0.7848	1.0000	0.1346
Distance . . .	0.0983	-0.1600	0.1993	0.0808	0.1545	-0.0910	0.0570	0.1346	1.0000

Verbal test 2

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0000	0.3270	0.0727	0.0000	0.1565	-0.1425	0.7309	-0.0379
2	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
3	0.3270	0.0000	1.0000	0.2543	0.0000	0.1061	0.0671	0.7520	0.1063
4	0.0727	0.0000	0.2543	1.0000	0.0000	0.2078	0.0829	0.4099	-0.0818
5	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
6	0.1565	0.0000	0.1061	0.2078	0.0000	1.0000	0.1206	0.3519	-0.0656
7	-0.1425	0.0000	0.0671	0.0829	0.0000	0.1206	1.0000	0.2880	0.0228
Ave.	0.7309	0.0000	0.7520	0.4099	0.0000	0.3519	0.2880	1.0000	0.0111
Distance . . .	-0.0379	0.0000	0.1063	-0.0818	0.0000	-0.0656	0.0228	0.0111	1.0000

Table B4 (Continued)

Total verbal

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0396	0.1562	0.1070	0.1139	0.0228	0.1395	0.6560	0.0623
2	0.0396	1.0000	0.0852	-0.1002	0.3474	-0.0040	0.1861	0.1937	-0.1378
3	0.1562	0.0852	1.0000	0.1716	0.0428	-0.0778	0.0860	0.5282	0.1485
4	0.1070	-0.1002	0.1716	1.0000	-0.0341	0.2933	0.5113	0.6057	0.0586
5	0.1139	0.3474	0.0428	-0.0341	1.0000	-0.0056	0.1841	0.2166	0.1260
6	0.0228	-0.0040	-0.0778	0.2933	-0.0056	1.0000	0.2047	0.2773	-0.0851
7	0.1395	0.1861	0.0860	0.5113	0.1841	0.2047	1.0000	0.6598	0.0624
Ave.	0.6560	0.1937	0.5282	0.6057	0.2166	0.2773	0.6598	1.0000	0.1094
Distance . . .	0.0623	-0.1378	0.1485	0.0586	0.1260	-0.0851	0.0624	0.1094	1.0000

Math test 1

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.2023	-0.1048	0.4450	0.1048	0.0	-0.2774	0.5973	-0.5697
2	0.2023	1.0000	-0.2757	0.5319	0.1654	0.0	0.2917	0.7297	-0.5622
3	-0.1048	-0.2757	1.0000	0.0551	-0.1429	0.0	0.0	-0.0125	0.0510
4	0.4450	0.5319	0.0551	1.0000	0.4962	0.0	0.2917	0.8554	-0.2727
5	0.1048	0.1654	-0.1429	0.4962	1.0000	0.0	0.3780	0.4884	-0.1615
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	-0.2774	0.2917	0.0	0.2917	0.3780	0.0	1.0000	0.3976	0.0370
Ave.	0.5973	0.7297	-0.0125	0.8554	0.4884	0.0	0.3976	1.0000	-0.5991
Distance . . .	-0.5697	-0.5622	0.0510	-0.2727	-0.1615	0.0	0.0370	-0.5991	1.0000

Table B4 (Continued)

Math test 2

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.5651	-0.0930	0.3269	0.5200	0.3231	0.5769	0.6853	-0.6091
2	0.5651	1.0000	0.2469	0.3916	0.1955	0.5488	0.5476	0.7521	-0.0131
3	-0.0930	0.2469	1.0000	0.4352	0.2045	0.1355	0.2613	0.4406	-0.0711
4	0.3269	0.3916	0.4352	1.0000	0.5067	0.6549	0.6133	0.8268	-0.3899
5	0.5200	0.1955	0.2045	0.5067	1.0000	0.1420	0.5380	0.6096	-0.6608
6	0.3231	0.5488	0.1355	0.6549	0.1420	1.0000	0.3915	0.6650	-0.0819
7	0.5769	0.5476	0.2613	0.6133	0.5380	0.3915	1.0000	0.8054	-0.3310
Ave.	0.6853	0.7521	0.4406	0.8268	0.6096	0.6655	0.8054	1.0000	-0.4411
Distance . . .	-0.6091	-0.0131	-0.0711	-0.3899	-0.6608	-0.0819	-0.3310	-0.4411	1.0000

Math test 3

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	-0.0354	-0.0147	0.1157	-0.3644	-0.0625	-0.1387	0.2822	-0.0219
2	-0.0354	1.0000	0.4167	0.0655	-0.4123	-0.0354	0.2550	0.5666	-0.0267
3	-0.0147	0.4167	1.0000	-0.4364	0.0000	0.2357	0.2451	0.6318	-0.1707
4	0.1157	0.0655	-0.4364	1.0000	-0.4499	-0.5401	-0.4708	-0.2264	-0.1185
5	-0.3644	-0.4123	0.0000	-0.4499	1.0000	0.7289	0.4043	0.1645	0.1751
6	-0.0625	-0.0354	0.2357	-0.5401	0.7289	1.0000	0.4507	0.5220	-0.0044
7	-0.1387	0.2550	0.2451	-0.4708	0.4043	0.4507	1.0000	0.6261	0.0552
Ave	0.2822	0.5666	0.6318	-0.2264	0.1645	0.5220	0.6261	1.0000	-0.0666
Distance . . .	-0.0219	-0.0267	-0.1707	-0.1185	0.1751	-0.0044	0.0552	-0.0666	1.0000

Table B4 (Continued)

Math test 4

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.0000	-0.2075	0.0299	0.3158	0.0526	0.0000	0.6227	0.1501
2	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	-0.0000
3	-0.2075	0.0000	1.0000	-0.1439	-0.2536	0.2075	0.0000	0.3616	0.1540
4	0.0299	0.0000	-0.1439	1.0000	-0.0299	-0.6272	0.0000	0.2219	-0.2577
5	0.3158	0.0000	-0.2536	-0.0299	1.0000	-0.0526	0.0000	0.5358	0.4803
6	0.0526	0.0000	0.2075	-0.6272	-0.0526	1.0000	0.0000	0.1014	0.1783
7	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	-0.0000
Ave.	0.6227	0.0000	0.3616	0.2219	0.5358	0.1014	0.0000	1.0000	0.3677
Distance . . .	0.1501	-0.0000	0.1540	-0.2577	0.4803	0.1783	-0.0000	0.3677	1.0000

Total math

	1	2	3	4	5	6	7	Ave.	Distance
1	1.0000	0.4284	-0.0336	0.3501	0.2010	0.2651	0.2125	0.6580	-0.4134
2	0.4284	1.0000	0.1038	0.4390	0.0706	0.4000	0.4216	0.7402	-0.1827
3	-0.0336	0.1038	1.0000	0.1625	-0.0366	0.1772	0.1878	0.3441	-0.0158
4	0.3501	0.4390	0.1625	1.0000	0.2524	0.3967	0.3444	0.7461	-0.3168
5	0.2010	0.0706	-0.0366	0.2524	1.0000	0.1684	0.3380	0.4065	-0.1669
6	0.2651	0.4000	0.1772	0.3967	0.1684	1.0000	0.3362	0.5834	-0.0344
7	0.2125	0.4216	0.1878	0.3444	0.3380	0.3362	1.0000	0.6274	-0.1419
Ave.	0.6580	0.7402	0.3441	0.7461	0.4065	0.5834	0.6274	1.0000	-0.3481
Distance . . .	-0.4134	-0.1827	-0.0158	-0.3168	-0.1669	-0.0344	-0.1419	-0.3481	1.0000

Table B5. Verbal Test 1, Question Data

Question	Delta		Distance from major axis	Rater	
	May 1976	May 1963		Mean	S.D.
1	9.0	8.1	0.22	3.29	0.70
2*	11.7	12.1	-0.70	2.14	0.64
3	13.0	11.5	0.64	3.00	0.00
4	11.7	11.4	-0.20	2.86	0.35
5*	12.8	11.8	0.29	2.29	0.88
6*	14.1	13.8	-0.20	2.43	0.73
7*	15.5	14.9	0.01	2.43	0.73
8*	15.6	15.3	-0.20	2.57	0.73
9	15.6	14.9	0.08	3.00	1.07
10**	15.8	16.0	-0.56	2.71	1.03
11*	9.2	8.3	0.22	2.57	0.73
12	10.7	10.9	-0.56	2.86	0.35
13*	12.1	11.6	-0.06	2.57	0.49
14	12.6	11.9	0.08	2.71	0.45
15*	13.7	12.2	0.64	2.57	0.73
16*	15.6	15.6	-0.42	2.00	0.93
17**	17.5	17.2	-0.20	2.86	0.83
18**	16.5	16.4	-0.34	2.14	0.99
19**	16.7	17.5	-0.98	2.29	0.88
20**	17.9	17.9	-0.41	2.14	0.99
21	7.6	8.5	-1.05	3.00	0.00
22**	5.5	6.0	-0.77	2.86	0.35
23	7.4	7.9	-0.77	2.86	0.35
24	10.0	10.6	-0.84	3.00	0.00
25*	11.2	9.8	0.57	2.57	0.49
26	12.5	11.1	0.57	2.86	0.64
27	11.6	12.6	-1.12	3.00	0.76
28	12.7	13.0	-0.63	2.71	0.88
29*	14.5	14.9	-0.70	2.57	0.73
30**	16.2	16.7	-0.77	3.00	1.07
31	11.0	10.1	0.22	2.71	0.45
32*	13.0	12.3	0.08	2.57	0.49
33	10.9	10.7	-0.28	2.86	0.35
34**	16.8	15.9	0.22	2.57	0.73
35*	12.5	12.0	-0.06	2.57	0.49
36	14.9	14.8	-0.34	2.71	0.70
37	15.4	14.3	0.36	2.86	0.35
38**	16.3	15.8	-0.06	2.71	0.45
39	11.5	11.5	-0.42	3.00	0.00
40	13.2	13.3	-0.49	3.00	0.00
41	11.0	8.1	1.63	3.14	0.35
42	11.1	9.0	1.07	3.29	0.70
43	11.7	10.2	0.64	3.14	0.35
44	11.2	10.9	-0.20	2.86	0.83

(Continued)

Table B5 (Continued)

Question	Delta May 1976	Delta May 1963	Distance from major axis	Rater Mean	Rater S.D.
45	14.1	13.4	0.08	2.86	0.35
46**	16.7	16.5	-0.27	2.14	0.64
47	11.4	10.4	0.29	3.00	0.53
48**	16.5	14.8	0.79	3.00	0.53
49	14.4	14.3	-0.34	2.86	0.64
50	11.5	10.7	0.15	3.14	0.83
51	9.9	9.2	0.08	3.00	0.00
52	10.9	9.9	0.29	3.00	0.53
53	12.3	11.1	0.43	3.29	0.70
54*	12.3	11.0	0.50	2.57	0.73
55*	12.9	12.6	-0.20	3.43	0.73
56	13.3	13.2	-0.35	2.86	0.35
57**	16.2	15.6	0.01	2.29	0.88
58*	15.5	15.7	-0.56	2.14	0.99
59	7.9	6.8	0.36	3.14	0.35
60	12.5	9.9	1.42	2.86	0.35
61	12.9	12.2	0.08	2.71	0.45
62*	12.9	13.1	-0.56	2.57	0.73
63*	14.6	15.0	-0.70	2.29	0.88
64	15.5	15.2	-0.20	2.71	0.45
65**	17.5	16.8	0.08	3.14	0.35
66**	16.4	17.5	-1.19	2.43	0.73
67	7.6	7.4	-0.28	2.86	0.35
68	8.3	8.1	-0.28	3.00	0.00
69	9.0	8.7	-0.21	3.14	0.35
70	11.4	10.9	-0.06	2.86	0.35
71	11.4	11.1	-0.20	2.71	0.45
72*	12.9	12.8	-0.35	2.43	1.05
73*	13.8	13.3	-0.06	2.43	1.05
74*	15.3	15.0	-0.20	2.29	0.88
75	15.0	13.1	0.93	3.00	0.00
76*	12.9	11.7	0.43	2.14	0.83
77*	13.1	12.5	0.01	2.43	0.73
78*	14.4	14.0	-0.13	2.29	0.70
79**	17.0	16.7	-0.20	2.29	0.70
80*	14.5	13.8	0.08	2.14	0.83
81	12.3	12.2	-0.35	3.00	0.53
82	13.7	13.4	-0.20	2.71	0.45
83	12.4	11.6	0.15	2.86	0.64
84	13.6	13.3	-0.20	2.86	0.64
85*	15.5	15.0	-0.06	2.57	0.90
86**	16.5	15.8	0.08	3.43	0.73
87*	11.3	11.4	-0.49	3.71	1.03
88	13.9	13.8	-0.35	3.14	0.83

(Continued)

Table B5 (Continued)

Question	Delta May 1976	Delta May 1963	Distance from major axis	Rater Mean	Rater S.D.
89	14.7	13.9	0.15	3.29	1.03
90*	11.0	11.1	-0.49	3.43	0.73
Delta mean	13.1	12.6	Rater mean	2.76	
Delta S.D.	2.6	2.7	Rater S.D.	0.35	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.
**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B6. Verbal Test 2, Question Data

Question	Delta May 1976	Delta Mar. 1973	Distance from major axis	Rater Mean	Rater S.D.
1	9.8	9.6	0.46	2.86	0.35
2	10.9	11.0	0.20	3.00	0.00
3	13.3	13.6	-0.03	3.00	0.53
4*	14.1	14.6	-0.21	2.57	0.73
5**	16.0	16.5	-0.28	2.71	0.45
6	8.1	9.0	-0.27	3.00	0.00
7	11.7	12.5	-0.33	3.14	0.83
8	14.8	14.9	0.06	2.86	0.35
9**	15.8	16.1	-0.12	2.86	0.35
10**	16.7	16.2	0.42	2.57	0.73
11	8.7	9.9	-0.51	2.71	0.45
12	10.2	12.0	-1.00	2.71	0.88
13	12.6	13.1	-0.15	3.00	0.53
14	15.5	15.6	0.03	3.14	0.35
15**	15.8	16.0	-0.05	2.86	0.83
16	10.0	10.4	0.02	3.14	0.35
17	11.5	12.0	-0.11	3.14	0.35
18	11.9	12.4	-0.12	2.86	0.35
19	14.4	14.4	0.15	2.71	0.70
20**	16.0	16.2	-0.06	2.71	0.35
21	9.3	9.8	-0.03	3.00	0.00
22	14.3	14.6	-0.07	2.71	0.45

(Continued)

Table B6 (Continued)

Question	Delta		Distance from major axis	Rater	
	May 1976	Mar. 1973		Mean	S.D.
23	15.7	15.0	0.60	2.86	0.64
24	14.0	14.4	-0.13	3.00	0.53
25**	17.5	17.5	0.03	2.43	0.73
26	9.2	9.9	-0.17	3.14	0.35
27	11.6	12.4	-0.33	3.00	0.53
28	13.8	13.6	0.31	2.86	0.35
29*	15.2	15.7	-0.25	2.43	0.90
30**	16.8	17.7	-0.60	2.71	0.45
31	10.4	10.2	0.44	2.86	0.83
32	14.3	14.4	0.08	2.71	0.70
33	10.2	10.8	-0.13	2.86	0.35
34	9.9	10.5	-0.12	3.00	0.53
35	15.2	15.3	0.04	3.00	0.53
36	12.8	12.9	0.13	2.71	0.70
37*	14.2	14.2	0.15	2.57	0.73
38*	14.8	15.0	-0.01	2.57	0.73
39**	16.0	16.2	-0.06	2.43	0.73
40*	14.1	14.7	-0.28	2.43	0.73
41	9.0	9.2	0.20	3.14	0.35
42	11.8	12.3	-0.12	3.14	0.35
43	12.5	12.6	0.14	3.14	0.35
44	13.8	13.7	0.24	3.14	0.35
45	11.1	11.4	0.05	2.86	0.35
46	11.9	12.3	-0.05	2.71	0.45
47	13.1	13.2	0.12	3.00	0.00
48*	13.3	13.4	0.11	2.57	0.73
49	10.8	11.1	0.06	3.00	0.00
50	12.0	12.6	-0.20	2.86	0.35
51**	6.0	7.4	-0.56	3.14	0.35
52*	8.3	9.0	-0.14	3.43	0.73
53	9.3	9.0	0.55	3.00	0.00
54	14.8	15.2	-0.16	2.71	0.88
55**	15.7	16.0	-0.12	2.57	0.49
56	14.8	14.8	0.13	3.00	0.53
57**	17.7	17.1	0.46	2.71	0.45
58**	16.5	17.3	-0.51	2.71	0.45
59	6.8	6.8	0.43	3.00	0.00
60	9.9	10.0	0.24	3.00	0.00
61	11.5	11.8	0.04	2.86	0.35
62	12.6	12.7	0.14	2.71	0.45
63	13.1	13.1	0.19	2.86	0.35
64	13.8	14.5	-0.34	2.86	0.35
65**	16.5	16.0	0.43	2.43	0.73
66**	17.1	17.7	-0.39	2.43	0.73

(Continued)

Table B6 (Continued)

Question	Delta		Distance from major axis	Rater	
	May 1976	Mar. 1973		Mean	S.D.
67	8.5	8.9	0.07	2.86	0.35
68	9.2	10.2	-0.39	2.86	0.35
69*	9.5	10.0	-0.04	2.43	0.49
70*	10.5	10.3	0.44	2.29	0.88
71	12.5	12.9	-0.07	2.71	0.45
72	12.3	13.0	-0.28	2.86	0.35
73**	16.4	16.1	0.29	2.57	0.73
74**	16.8	17.0	-0.09	2.57	0.73
75**	17.7	18.2	-0.34	2.29	0.88
76*	11.5	11.6	0.18	2.57	0.73
77*	11.0	11.3	0.05	2.57	0.73
78**	16.4	16.6	-0.07	2.57	0.49
79	14.2	14.4	0.01	2.71	0.45
80*	13.0	13.4	-0.09	2.43	0.73
81	13.9	14.1	0.02	3.00	0.53
82	13.9	14.2	-0.05	2.71	0.45
83	12.2	12.3	0.15	3.14	0.35
84	14.3	15.1	-0.43	3.00	0.53
85	12.2	12.6	-0.06	2.71	0.88
86	13.5	13.5	0.18	2.71	0.45
87	13.7	13.7	0.17	2.71	0.45
88*	11.6	11.7	0.18	2.57	0.73
89	14.7	14.8	0.06	2.71	0.45
90	13.7	14.0	-0.05	2.71	0.45
Delta mean	12.9	13.2	Rater mean	2.80	
Delta S.D.	2.7	2.5	Rater S.D.	0.23	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.
**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B7. Mathematics Test 1, Question Data

Question	Delta		Distance from major axis	Rater	
	Jan. 1971	May 1962		Mean	S.D.
1	8.9	8.3	-0.09	2.86	0.35
2**	7.4	6.0	0.50	3.14	0.35
3	11.1	10.3	0.02	3.00	0.53
4	9.9	9.9	-0.53	3.29	0.45
5	9.0	8.1	0.12	3.14	0.64
6	13.1	11.9	0.28	2.86	0.35
7*	11.9	11.8	-0.49	3.57	0.49
8	13.2	11.9	0.35	3.00	0.53
9	14.0	12.1	0.76	2.86	0.35
10*	12.9	11.9	0.13	2.43	0.49
11	13.5	12.6	0.05	3.00	0.53
12*	15.5	14.5	0.10	2.57	0.49
13	14.5	14.7	-0.75	3.29	0.70
14	13.8	13.9	-0.66	3.14	0.83
15	14.8	13.4	0.39	2.86	0.35
16	14.5	13.6	0.04	3.14	0.35
17**	16.5	16.7	-0.78	2.29	0.70
18	15.4	14.5	0.03	3.14	0.35
19**	16.0	15.5	-0.27	3.14	0.99
20**	16.8	15.8	0.08	2.57	0.90
21**	16.2	16.9	-1.13	2.86	0.64
22**	16.4	16.7	-0.85	2.71	0.70
23**	16.1	16.5	-0.91	2.57	0.73
24**	15.4	16.0	-1.05	3.14	0.64
25**	17.4	17.5	-0.72	3.00	0.00
Delta mean	13.8	13.2	Rater mean	2.94	
Delta S.D.	2.7	3.0	Rater S.D.	0.29	

*Questions excluded from delta plots because rater mean is less than 2.7 or greater than 3.3.
**Questions excluded from delta plots and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B8. Mathematics Test 2, Question Data

Question	Delta		Distance from major axis	Rater	
	Dec. 1973	Jan. 1965		Mean	S.D.
1	6.8	7.1	-0.92	3.14	0.35
2*	10.4	8.5	0.89	2.29	0.45
3	9.5	9.0	-0.11	3.29	0.45
4	10.5	9.8	0.12	3.14	0.64
5*	8.5	9.3	-1.07	3.71	0.70
6	10.5	9.2	0.51	3.00	0.53
7	11.2	10.9	-0.06	2.86	0.64
8	9.8	7.9	0.83	3.00	0.53
9*	11.0	11.0	-0.28	3.57	0.49
10*	12.4	11.8	0.26	3.43	0.90
11*	10.6	11.5	-0.91	4.43	0.49
12	13.0	12.6	0.19	3.14	0.35
13	13.2	13.8	-0.44	2.86	0.64
14*	13.2	13.9	-0.50	3.71	0.35
15	14.5	15.3	-0.43	3.14	0.35
16*	14.2	14.1	0.12	3.57	0.49
17*	13.2	13.1	0.02	4.14	0.64
18*	15.4	15.4	0.19	4.14	0.64
19*	13.6	13.7	-0.07	3.57	0.49
20	14.7	14.4	0.31	3.29	0.70
21*	15.1	15.0	0.22	3.57	0.49
22**	16.1	15.1	0.91	3.14	0.35
23*	14.8	15.1	-0.07	3.57	0.73
24**	15.9	15.5	0.50	3.14	0.64
25**	16.8	17.1	0.14	3.43	0.49
Delta mean	12.6	12.4	Rater mean	3.37	
Delta S.D.	2.6	2.8	Rater S.D.	0.45	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.
**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B9. Mathematics Test 3, Question Data

Question	Delta		Distance from major axis	Rater	
	Nov. 1975	Apr. 1971		Mean	S.D.
1	9.3	7.9	0.69	3.00	0.00
2	8.7	8.1	0.12	3.00	0.00
3	9.5	9.5	-0.27	3.00	0.00
4	13.0	12.3	0.31	3.29	0.45
5	9.8	10.3	-0.61	3.14	0.35
6	10.8	10.7	-0.16	2.86	0.35
7	11.5	11.1	0.06	3.29	0.70
8	9.2	9.2	-0.28	3.29	0.70
9	12.3	12.0	0.02	3.14	0.35
10	13.0	12.9	-0.10	3.14	0.35
11	13.5	13.3	-0.02	3.29	0.45
12*	14.0	13.8	-0.01	3.43	0.73
13	13.8	13.6	-0.01	3.00	0.00
14	11.9	11.4	0.14	3.00	0.53
15	14.7	14.3	0.15	3.00	0.00
16	11.6	11.2	0.07	3.14	0.64
17**	15.9	15.8	-0.02	3.00	0.00
18**	16.2	15.7	0.26	3.14	0.64
19	14.8	14.8	-0.12	3.00	0.00
20**	15.4	16.0	-0.52	3.00	0.00
21**	15.7	16.1	-0.37	3.00	0.53
22**	16.1	16.2	-0.15	3.29	0.45
23**	16.2	16.3	-0.15	2.86	0.64
24**	16.8	15.8	0.63	2.71	0.45
Delta mean	13.1	12.8	Rater mean	3.06	
Delta S.D.	2.5	2.6	Rater S.D.	0.19	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Table B10. Mathematics Test 4, Question Data

Question	Delta		Distance from major axis	Rater	
	Apr. 1976	Mar. 1972		Mean	S.D.
1	8.1	8.3	-0.10	3.14	0.35
2	8.7	8.3	0.32	3.14	0.35
3	8.7	9.0	-0.17	3.00	0.00
4	10.2	10.0	0.17	2.86	0.35
5	10.5	10.7	-0.11	3.14	0.35
6	10.0	10.2	-0.11	3.14	0.35
7	11.6	12.0	-0.26	3.29	0.45
8	10.3	10.8	-0.32	3.00	0.00
9	13.5	13.1	0.30	3.29	0.45
10	12.8	12.2	0.45	3.14	0.35
11	14.9	14.7	0.16	3.14	0.35
12	13.3	13.1	0.16	3.14	0.35
13	14.1	13.8	0.23	3.14	0.64
14	14.5	15.4	-0.62	2.86	0.64
15	12.6	13.9	-0.19	3.00	0.00
16	14.7	13.9	0.58	3.14	0.35
17	14.9	15.0	-0.06	3.00	0.00
18	14.7	15.4	-0.48	3.14	0.35
19**	14.8	16.0	-0.83	2.86	0.35
20	15.4	15.3	0.08	3.14	0.35
21**	15.8	16.1	-0.20	3.14	0.35
22**	18.2	18.3	-0.07	2.71	0.45
23	13.9	14.0	-0.05	3.00	0.53
24**	16.9	16.2	0.51	2.86	0.35
Delta mean	13.0	13.1	Rater mean	3.05	
Delta S.D.	2.7	2.7	Rater S.D.	0.15	

*Questions excluded from delta plot because rater mean is less than 2.7 or greater than 3.3.

**Questions excluded from delta plot and other analyses because delta was less than 6.4 or greater than 15.7 at at least one administration.

Figure B1. Relationship between Delta and Percent Correct

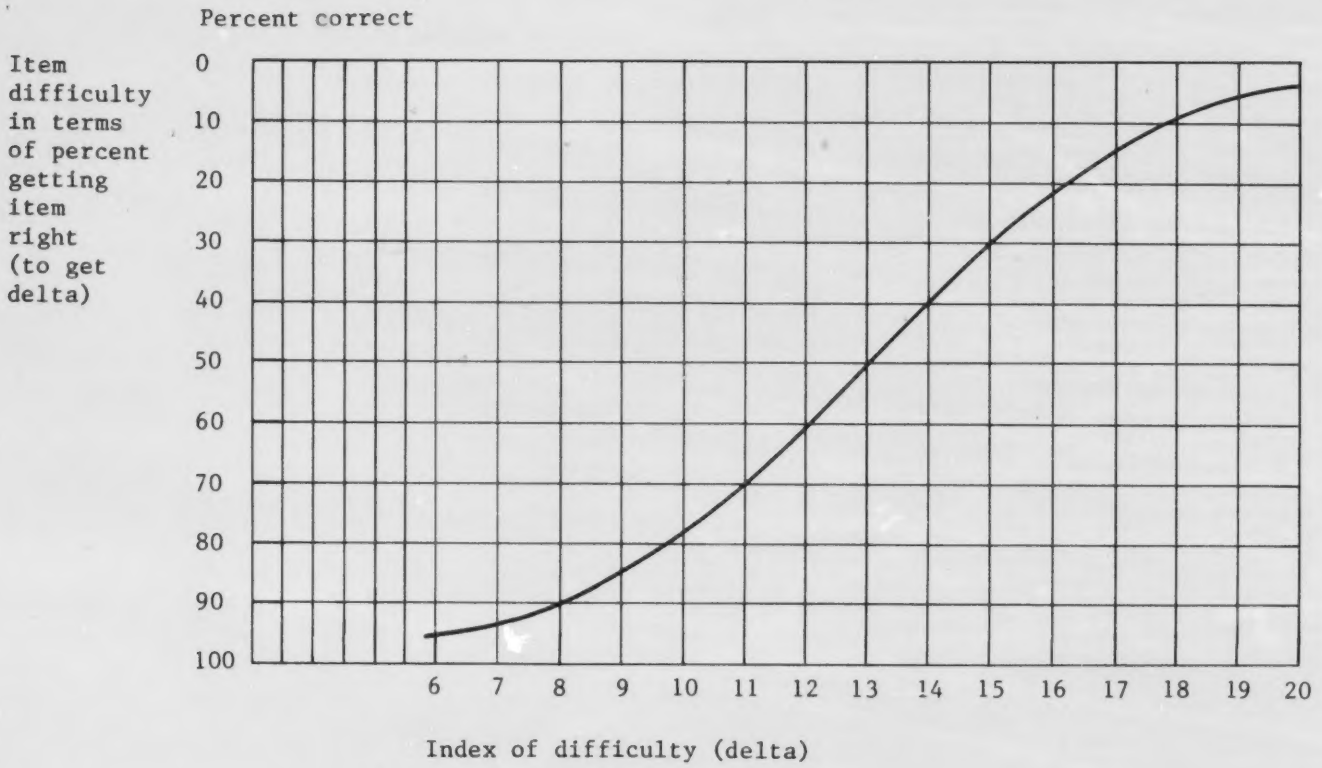
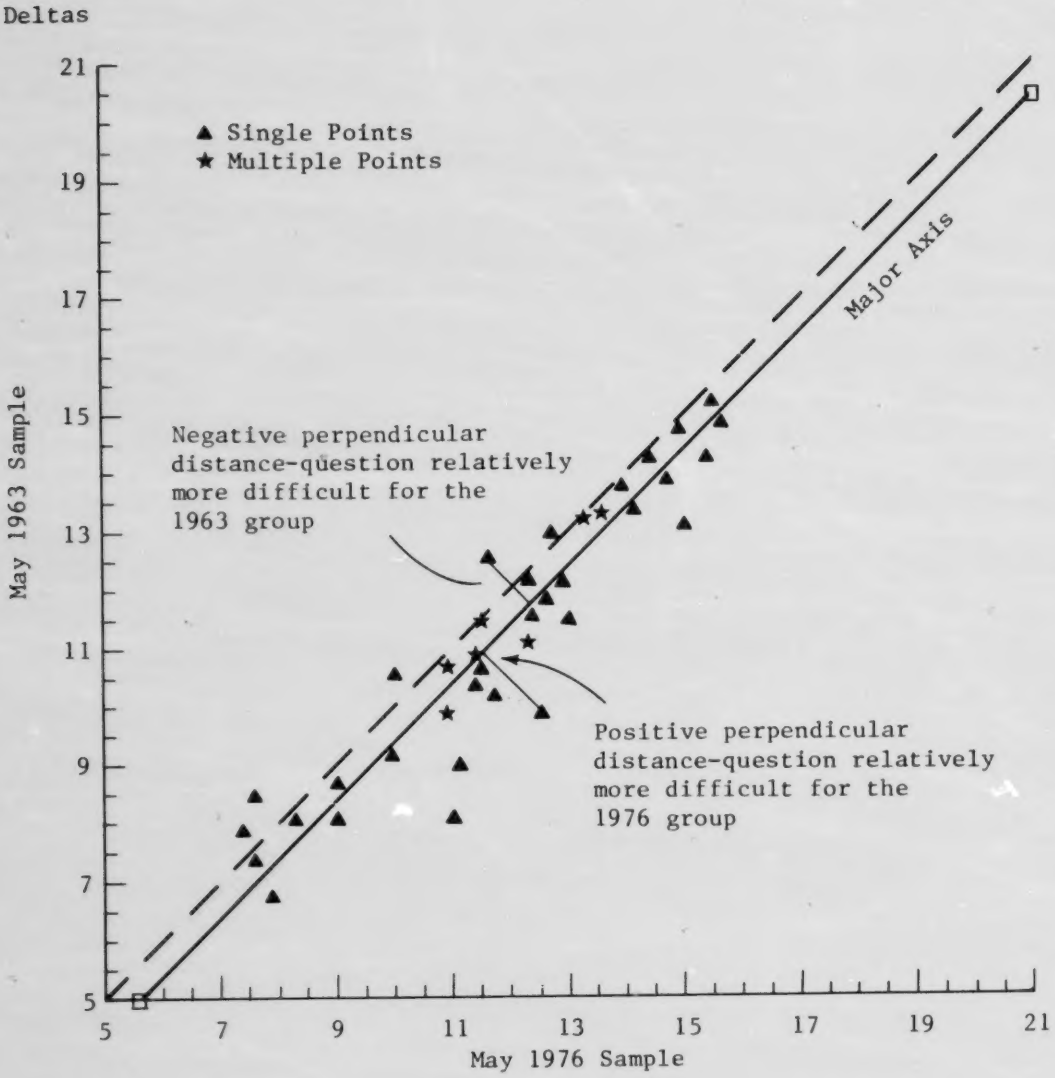


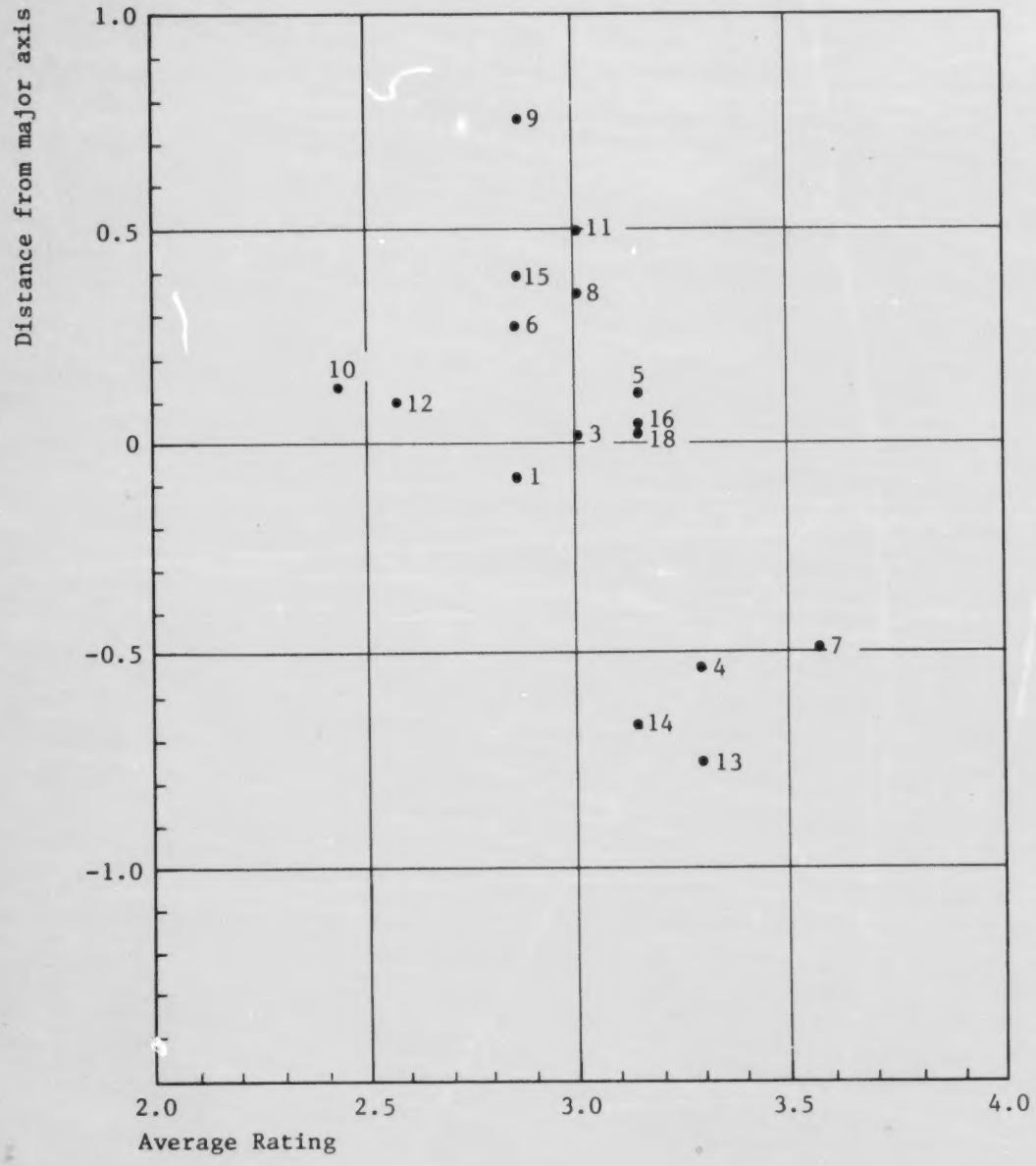
Figure B2. Delta Plot for Verbal Test 1



Plot TIS	Test	Form	N
50015001	SAT-V	LSA33	046
50025002	SAT-V	LSA33	046

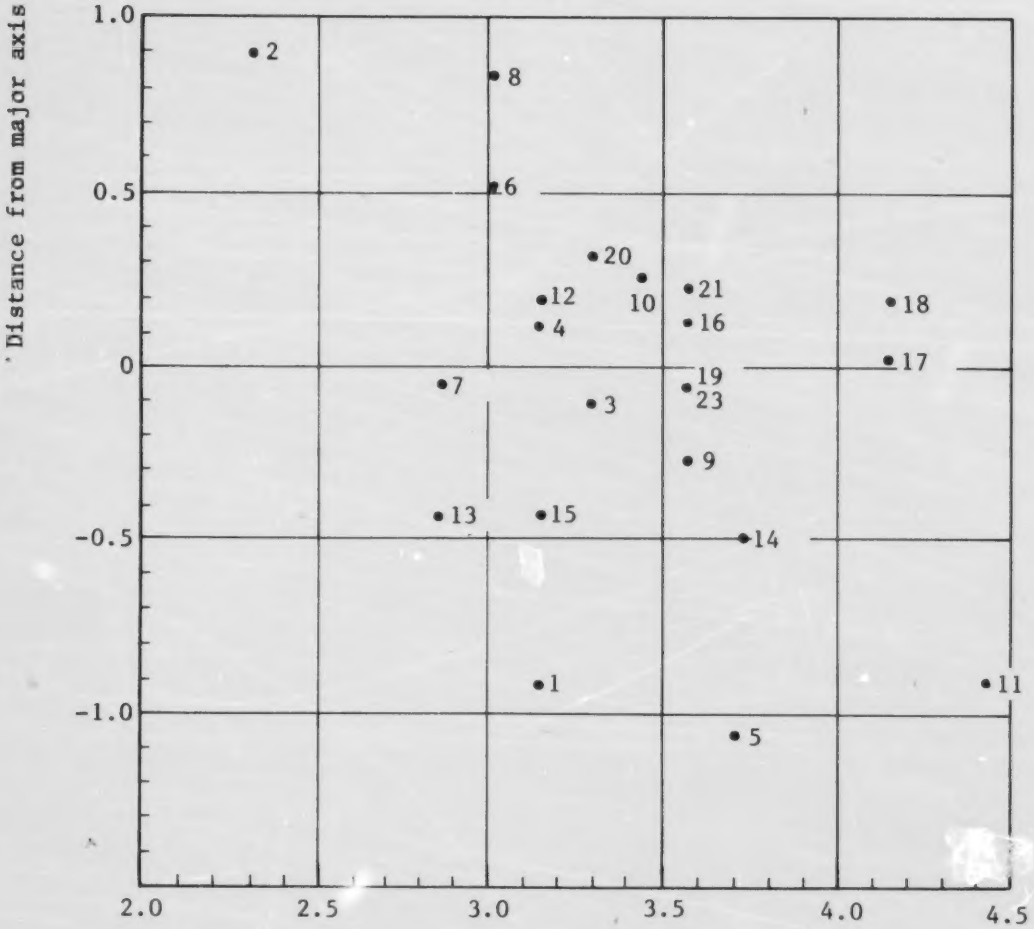
May 1976 Sample
May 1963 Sample

Figure B3. Scatter Plot for Math Test 1
Average Rating vs Distance from Major Axis



Numbers refer to the item number in Test 1.

Figure B4. Scatter Plot for Math Test 2
Average Rating vs Distance from Major Axis



Numbers refer to the item number in Test 2.

APPENDIX C

QUESTIONS IDENTIFIED AS EXTREME OUTLIERS

Table C1

The Five Verbal Questions with the Highest Rater Means (High rater mean and negative perpendicular distance both indicate a question that should be easier for the more recent group.)

Question 55, Test 1	Rater Mean	3.43
	Perpendicular Distance	-0.20

55. Some firms have so ------- their line of products that, when the seasonal demand for one product slackens, the demand for some other one increases.

(A) consolidated
(B) refined
* (C) diversified
(D) equalized
(E) standardized

Comments on Question 55:

1. Consumer economics vocabulary more in use today in social studies.

* Denotes correct answer

Passage for Questions 86-90

The word cybernetics is from the Greek for "steersman," from the same root as the word "governor." It is Wiener's coinage for the new science of "control and communication in the animal and the machine." The element of control is that exemplified in the mechanical device known as a governor, operating on the fundamental principle of the feedback: the working of the machine at a given time is itself a stimulus for the modification of the future course of its working. Communication enters into cybernetics by way of an unorthodox and ingeniously precise concept of "information" as the degree to which the impulses entering a machine reduce the uncertainty among the set of alternatives available to it. Thus if a machine contains a relay which has two possible positions, an impulse which puts the relay into one of these two has conveyed to the machine exactly one "bit" of information. All communication can be regarded as made up of such bits, as is suggested by the dots and dashes of the Morse code, for example. Modern machines actually use a binary arithmetic; i.e., a notation in which all numbers are expressed in terms of powers of 2 (rather than 10), and are therefore written as a series of "zeros" and "ones." One bit of information is conveyed by each choice of either a zero or a one—the word "bit" being an abbreviation for "binary digit."

Working with these concepts of communication and control, cybernetics becomes relevant to the study of man because human behavior is paralleled in many respects by the communication machines. This parallel is no mere metaphor, but consists in a real similarity between the machine processes and those of human behavior. The Darwinian continuity between man and the rest of nature has now been carried to completion: man's rationality marks only a difference of degree from other animals and, fundamentally, no difference at all from the machine. For modern computers are essentially logical machines: they are designed to confront propositions and to draw from them their logical conclusions. With communication and control as the key, a similarity can also be traced between an individual (whether human or mechanical) and a society (again, whether men or machines in a well-designed factory). The metaphors of Plato and Hobbes can now be given a literalist interpretation.

Thus cybernetics bears on the study of human behavior in a variety of ways: most directly, by way of neurology and physiological psychology; and by simple extension, to an improved understanding of functional mental disorders, which Wiener finds to be primarily diseases of memory, thus arriving at Freudian conclusions by a totally different route. Of particular interest are the implications, definitely present though not explicitly drawn, for such classical philosophical puzzles as those concerning free will and the mind-body relation. The analysis of mind and individual personality as a structure of certain information processes renders obsolete not only the "mind substance" of the idealist, but mechanistic materialism as well. Mind is a patterning of information and not spirit, matter, or energy.

Question 86, Test 1

Rater Mean 3.43
Perpendicular Distance 0.08

86. According to the passage, which of the following concepts are fundamental to the science of cybernetics?

(A) Man and machines
*(B) Information and regulation
(C) Computers and logic
(D) Binary digits and bits
(E) Impulses and probabilities

Question 87

Rater Mean 3.71
Perpendicular Distance -0.49

87. According to the explanation in the passage, which of the following is a binary digit?

(A) $\frac{1}{2}$ *(B) 1 (C) 2 (D) 4 (E) 23

Question 90

Rater Mean 3.43
Perpendicular Distance -0.49

90. On the basis of the passage, how many "bits" are in 01001?

(A) 2 (B) 3 (C) 4 *(D) 5 (E) 1001

Comments on Questions 86, 87, and 90:

1. Computer related terms are more familiar.
2. Difficult passage in vocabulary, concepts, and demands on analogy.
3. Today's students have more experience with the computer and related vocabulary.
4. Timeliness of content
5. Question 90 is essentially a mathematical question--computers are "in" for many students.

Question 52, Test 2

Rater Mean 3.43
Perpendicular Distance -0.14

52. Engineers have long ------- that magnetic tape would ------- standard phonograph records, since it wears longer and is difficult to scratch.

*(A) predicted..replace (B) warned..destroy
(C) admitted..ruin (D) lamented..surpass
(E) charged..outsell

Comments on Question 52

1. A societal difference--a current technological awareness.
2. Students have much tape and little record experience.

Table C2

The Five Verbal Questions with the Lowest Rater Means (Lower mean indicates the question should be more difficult for the more recent group)

Question 2, Test 1

Rater Mean 2.14
Perpendicular Distance -0.70

2. Although smallpox has been and still is a severe and dreadful disease, there is some evidence that present-day smallpox has lost part of its -------.

(A) immunity ★(B) virulence (C) effectiveness
(D) symptoms (E) activity

Comments on Question 2:

1. Vocabulary--"virulence" rarely used today--few epidemics written about in news.
2. Difficult vocabulary.

Question 18, Test 1

Rater Mean 2.14
Perpendicular Distance -0.34

18. EXCULPATE: (A) praise (B) promote
★(C) prove guilty (D) welcome cordially
(E) detract

Comments on Question 18:

1. A rarely used term today.
2. Difficult vocabulary
3. No longer in common usage

Question 58, Test 1

Rater Mean 2.14
Perpendicular Distance -0.56

58. Like an apéritif, the book stimulates without -------, the -------.

(A) reaching..climax
(B) calming..nerves
★(C) satisfying..appetite
(D) exciting..emotions
(E) exhausting..possibilities

Comments on Question 58:

1. Many do not know foreign words.
2. Fewer foreign language electives.
3. Note choice (C).

Passage for questions 76 and 80, Test 1

I feel not in me those sordid and unchristian desires of my profession: I do not secretly implore and wish for plagues, rejoice at famines, consult almanacs in expectation of malignant aspects, fatal conjunctions, and eclipses. I rejoice not at unwholesome springs nor unseasonable winters: my prayer goes with the husbandman's. Let me be sick myself if sometimes the malady of my patient be not a disease unto me. I desire rather to cure his infirmities than my own necessities. I am not only ashamed but heartily sorry that, besides death, there are diseases incurable: yet not for my own sake, or that they be beyond my art, but for the general cause and sake of humanity, whose common cause I apprehend as my own. And to speak more generally, those three noble professions which all civil commonwealths do honor are raised upon the fall of Adam and are not anyway exempt from their infirmities; there are not only diseases incurable in Physick, but cases indissolvable in Laws, vices incorrigible in Divinity. I can cure the gout or stone in some sooner than Divinity pride or avarice in others. I boast nothing, but plainly say we all labor against our own cure; for death is the cure of all diseases. There is no universal remedy I know but this, which, though nauseous to queasy stomachs, yet to prepared appetites is nectar and a pleasant potion of immortality.

76. The author indirectly criticizes some of his colleagues by implying that they

(A) put their faith in universal remedies
(B) are proud and avaricious
*(C) care more for gain than for human welfare
(D) allow superstition to influence their judgments
(E) fail to acknowledge the incurability of some diseases

Rater Mean 2.14
Perpendicular Distance 0.43

---

80. The sentence beginning "I can cure the gout or stone . . ." (lines 22-24) might be clarified by inserting

(A) commas before and after the word "pride"
(B) the words "I can cure" before the word "Divinity"
(C) the words "I can be cured of" before the word "Divinity"
*(D) the words "can cure" after the word "Divinity"
(E) the words "can be cured of" after the word "Divinity"

Rater Mean 2.14
Perpendicular Distance 0.08

Comments on Questions 76 and 80:

1. Close analytic reading of difficult prose has been neglected or abandoned in the many secondary school curriculums thereby making such questions more difficult.
2. Archaic usage and vocabulary
3. Less concern/experience with pre-20th century literature, hence with archaic diction.
4. Task archaic--divine Physick indeed.
5. Obscure construction in question 80.

Table C3

The Five Verbal Questions with the Greatest (Negative) Perpendicular Distance
(These questions were relatively easier for more recent groups)

Question 19, Test 1

Rater Mean 2.29
Perpendicular Distance -0.98

19. ESULLIENCE: *(A) apathy (B) ignominy
(C) antithesis (D) protectiveness
(E) frustration

Comments on Question 19:

1. Difficult vocabulary
2. Uncommon in jargon

Question 21, Test 1

Rater Mean 3.00
Perpendicular Distance -1.05

21. SUFFOCATE: OXYGEN:: (A) electrocute: current
*(B) starve: food (C) write: pencil
(D) burn: paper (E) murder: poison

No comments

Question 27, Test 1

Rater Mean 3.00
Perpendicular Distance -1.12

27. ANALGESIC: PAIN:: (A) submission: authority
*(B) silencer: noise (C) monotony: task
(D) pride: humiliation (E) funeral: death

Comments on Question 27:

1. Formerly analgesic was a common term used in commercials for aspirin--not today.
2. More concern in early 60's with sciences.

Question 66, Test 1

Rater Mean 2.43
Perpendicular Distance -1.19

66. HOYDEN: (A) pleasant companion
(B) generous friend
(C) old woman
(D) aristocratic youth
* (E) well-mannered girl

Comments on Question 66:

1. Archaic terminology
2. Difficult vocabulary
3. "Hoyden" obsolete

Question 12, Test 2

Rater Mean 2.71
Perpendicular Distance -1.00

12. HOMESTRETCH : RACE :: *(A) finale : opera
(B) goal : contest (C) boundary : journey
(D) terminal : station (E) platform : campaign

No comments

Table C4

The Five Verbal Questions with the Greatest (Positive) Perpendicular Distance
(These questions were relatively more difficult for the more recent group)

Passage for Questions 41–42, Test 1

The ways in which people undertake the waging of war depend not only upon the tools they possess but also upon their social organization and motivations. The chief reasons for war among primitive peoples are not economic. Real or fancied grudges, the desire for revenge, and extroverted antipathy come first. Reciprocity, both positive and negative, is the basis of all social relations. In 1933 an old Comanche gave voice to this in the expressed desire to scalp a young anthropologist "because he looks like a German," and no one had brought old Mumseka a German scalp for the life of his grandson who had fallen in France in 1918.

Religious and magical notions probably rank next after revenge drives. All head-hunting is tied up with supernaturalism and with the belief that dead men's power can be taken with their heads. Polynesian wars were frequently undertaken to obtain sacrificial victims. Ashantis and other West Africans kept prisoners for sacrifice. The bloody religion of the Aztecs called for thousands of war prisoners for human sacrifice to the gods—a fact that proved their undoing when they came to grips with Cortez. The Aztecs fought to take prisoners; the Spaniards fought to kill. They baptized only survivors.

The seeking of status and excitement is another important motivation for war. Pueblo Indians allow no special honor to warriors, but Plains Indians made the measurement of a man his war record. Few are the tribes that do not honor the warrior. Outright conquest was undertaken by Inca, Aztec, and many African kings as a means of enhancing personal glory and honor.

In the period of disillusionment following the First World War, it became the fashion to maintain that all wars spring from sordid economic greed. Self-enrichment is a mighty, driving power. War can serve well as a means of depriving other peoples of access to natural resources. Or it may be used as an instrument of subjugation for the purposes of callous exploitation. Whether it is so used or not depends upon various factors that are culturally determined.

War has become the deadly issue of the twentieth century. It now threatens the very existence of the societies that have brought it to its high peak of destructive perfection. After all, no culture can long nourish a trait that is actively and effectively self-destructive. But we are attempting just this, and it is our tragedy that, seeing where we are heading, we seem unable to alter our course

41. Which of the following best summarizes and epitomizes the Comanche concept of justice as exemplified by Mumseka?

(A) Do unto others as you would they would do unto you.
(B) Though justice moves slowly, it seldom fails to overtake the wicked.
(C) Let justice be done though the heavens fall.
★(D) An eye for an eye, and a tooth for a tooth.
(E) He injures the good who spares the bad.

Rater Mean	3.14
Perpendicular Distance	1.63

42. In those tribes where warriors were looked upon with favor, what was the principal motivation for individual participation in wars?

Rater Mean 3.29
Perpendicular Distance 1.07

(A) To improve the individual's economic situation
(B) To obtain sacrificial victims
*(C) To attain fame and recognition
(D) To propagate the tribal religion
(E) To avenge the tribal honor

Comments on Questions 41 and 42:

1. If the curriculum change in the past 13 years in social studies has had any effect in making anthropological content and concepts more familiar, today's students SHOULD be more comfortable with this passage.

Passage for question 48, Test 1

The increasingly timid use of the death penalty makes nonsense of the main arguments for retaining it: that it is a necessary form of retribution, an expression of condemnation of a certain type of crime, and a necessary deterrent against this same crime. Clearly, if some first degree murders are punished by death while others are not, society is expressing its condemnation only of particular murderers rather than of first degree murder as such. Hence the fine line between retribution and vengeance disappears.

Historically, capital punishment belongs to a penal system based on violence; the morality which fostered it has long been in retreat before humanitarian and scientific advances. Modern penology seeks to replace the concept of "the punishment to fit the crime" with a new notion: "the treatment to fit the criminal." The interests of society must, of course, be placed before the rights of the individual; no judge should permit dangerous criminals to remain at large. But society's interests would be adequately protected if such criminals were treated before committing a capital crime!

This could happen under a curative penal system. Murder is seldom the first crime committed, and a competent diagnostician usually can detect homicidal tendencies before these erupt. But little improvement can be expected so long as an archaic legal test of sanity—the ability to distinguish between right and wrong—holds abnormal people fully responsible to the law. The practice of punishing these helpless people does more than violate the rights of the individual: it threatens public safety. For, while punishment has no effect on them, its infliction means that society cannot be permanently safeguarded from them until they commit a crime so great that the legal sentence is life imprisonment or death. The law of criminal responsibility must be reformed if the problem that the abnormal offender represents is ever to be solved by curative means. And this is a reform, vital to modern penological principles, that the death penalty and other purely punitive symbols are holding back.

48. According to the passage, why have the instances of punishment by death become less frequent in recent years?

(A) Laws providing for capital punishment have been modified.
(B) Other forms of punishment have been proven to be more humane.
* (C) Those passing sentence have chosen to give other penalties.
(D) Deterrent influences have made capital crimes less frequent.
(E) Capital punishment has little effect in preventing crime.

Rater Mean	3.00
Perpendicular Distance	0.79

Comments on Question 48:

1. In the past 13 years criminology/sociology vocabulary and concepts have become a larger part of the secondary curriculum.
2. Recent concerns about death penalty may cause students to become concerned by opening statement. "Increasingly timid use" suggests there has not been recent publicity about re-instatement of death penalty in California. Some articles see the re-instatement as a backward step, not "increasingly timid use..."

Question 60, Test 1

60. SECEDE: (A) advance (B) promulgate
*(C) join (D) test (E) veto

Rater Mean	2.86
Perpendicular Distance	1.42

Comments on Question 60:

1. Harder because of absence of "civil war" era in conceptual social studies.

Question 75, Test 1

75. RABBIT : SPEED:: (A) muskrat : water
*(B) bee : industry (C) beaver : dam
(D) fox : evil (E) cat : worship

Rater Mean	3.00
Perpendicular Distance	0.93

No comment

Table C5

The Five Mathematical Questions with the Highest Rater Means (High mean indicates question should be easier for the more recent group)

Question 5, Test 2

Rater Mean 3.71
Perpendicular Distance -1.07

5. If $x - 0 = y$, then

(A) $x > y$ *(B) $x = y$ (C) $x < y$

(D) $\frac{y}{x} = 0$ (E) $\frac{x}{y} = 0$

Comments on Question 5:

1. Real number properties emphasized recently at all levels of instruction.
2. Between the two times of administration there was a great deal of emphasis on the properties of 0.

Question 11, Test 2

Rater Mean 4.43
Perpendicular Distance -0.91

11. The symbol $\bullet$ is an operation of arithmetic, a and b are real numbers, and $a \neq 0$. If $a \bullet b = b \bullet a$ and $a \bullet 0 = a$, then operation $\bullet$ is

*(A) addition
(B) subtraction
(C) multiplication
(D) division
(E) either addition or multiplication

Comments on Question 11:

1. 1973 group has had more experience with abstract idea of binary operations.
2. Now included in curriculum.
3. The modern math student would feel comfortable with * operation.
4. Real number properties emphasized recently at all levels of instruction.
5. Direct influence of emphasis in "new Math".

Question 14, Test 2

Rater Mean 3.71
Perpendicular Distance -.50

14. If the product of x and y is divisible by 12 and if y is an odd integer, then

(A) x is divisible by 3
*(B) x is divisible by 4
(C) x is divisible by 6
(D) y is divisible by 6
(E) y is divisible by 2

Comments on Question 14:

1. Number theory used more motivationally in recent years.
2. More work in the curriculum on number theory.

Questions 17-18, Test 2

Questions 17-18 refer to the following definitions.

The notation $x|y$ shall be defined as follows: $x|y = x + y$ if x and y are both odd or both even; $x|y$ = positive difference of x and y if one is odd and one is even. Thus, $1|2 = 1$; $2|4 = 6$; and $3|5 = 8$.

17. $\frac{1|2}{3|4} =$

Rater Mean 4.14
Perpendicular Distance 0.02

(A) 0 (B) $\frac{1}{7}$ (C) $\frac{3}{7}$ *(D) 1 (E) 3

---

18. If x is positive and $(2x + 1)|(4x + 4) = 9$, then $x =$

Rater Mean 4.14
Perpendicular Distance 0.19

(A) $\frac{2}{3}$

(B) 1

(C) 2

(D) $\frac{7}{3}$

*(E) 3

Comments on Questions 17-18:

1. 1973 group has had more exposure to abstract ideas and binary operations.
2. Definitions of operations are taught today.
3. More emphasis in the curriculum on interpreting notation.

Table C6

The Five Mathematical Questions with the Lowest Rater Means (Low mean indicates question should be more difficult for the more recent group)

Question 10, Test 1

Rater Mean 2.43
Perpendicular Distance 0.13

10. A fisherman caught 10 fish, each weighing 2 pounds, with 8 ounces of bait. At this rate, how many pounds of fish could he expect to catch using 2 pounds of bait?

(A) 20 (B) 40 (C) 60 *(D) 80 (E) 100

Comments on Question 10:

1. Ratio covered at the end of the school year.
2. Less practice in "new math" on these kinds of applications.
3. Go metric.

Question 12, Test 1

Rater Mean 2.57
Perpendicular Distance 0.10

12. If p is the price, in cents, of each apple, how many apples can be purchased for d dollars?

(A) $\frac{d}{p}$ (B) $\frac{p}{100d}$ (C) $\frac{d}{100p}$

(D) $\frac{100p}{d}$ *(E) $\frac{100d}{p}$

Comments on Question 12:

1. Ratio usually covered at the end of the school year.
2. Less practice in the "new math" on these kinds of applications.

Question 17, Test 1

Rater Mean 2.29
Perpendicular Distance -0.78

17. If gasoline costs 28.9 cents a gallon for regular and 31.9 cents a gallon for high-test, what is the difference, in gallons, between the amount of regular and the amount of high-test which $3.00 will buy?

(A) .28 (B) .31 (C) .90 *(D) .98 (E) 3

Comments on Question 17:

1. Problem not closely related to curriculum.
2. This problem would test quite a different skill if the calculator were used. Most students would do this by computing. Where can I buy gas at these prices?

Question 20, Test 1

Rater Mean 2.57
Perpendicular Distance 0.08

20. If 30 per cent of X is 40 per cent of Y, then X is what per cent of Y ?

(A) $33\frac{1}{3}$ (B) 70 (C) 75 (D) 120 *(E) $133\frac{1}{3}$

Comments on Question 20:

1. Less emphasis on per cent problems.
2. Ratio usually covered at the end of the school year.
3. Seem to do less with per cent.

Question 2, Test 2

Rater Mean 2.29
Perpendicular Distance 0.89

2. The list price of a certain book is $3.95. If it is sold at a 20 per cent discount, how much does the buyer save?

(A) $0.20
*(B) $0.79
(C) $0.95
(D) $1.98
(E) $3.16

Comments on Question 2:

1 Less work on these types of problems.

Table C7

The Five Mathematical Questions with the Greatest (Negative) Perpendicular Distances (These questions were relatively easier for more recent groups)

Question 21, Test 1

Rater Mean 2.86
Perpendicular Distance -1.13

21. An instrument dial is marked off in four numbered points, each 90 degrees apart. The pointer on the dial turns at the rate of 30 degrees per second from 0 to 2. After $2\frac{1}{2}$ seconds, what is the ratio of the distance already covered to the distance remaining to point number 2 ?

(A) $\frac{1}{3}$ (B) $\frac{4}{7}$ *(C) $\frac{5}{7}$ (D) $\frac{2}{3}$ (E) $\frac{5}{6}$

Comments on Question 21:

1. The wording is ambiguous. Supply a picture with the 4 numbered points.

Question 24, Test 1

Rater Mean 3.14
Perpendicular Distance -1.05

24. A model-M truck carries between 4 and 5 tons of bricks and a model-N truck carries between 5 and 7 tons of bricks. What is the smallest possible difference, in tons, between a load on 3 model-M trucks and a load on 2 model-N trucks?

*(A) 0 (B) 1 (C) 2 (D) 3 (E) 5

Comments on Question 24:

1. Inequalities given more study.

Question 1, Test 2

Rater Mean 3.14
Perpendicular Distance -0.92

1. If x times x is 25, then x times x times x is

(A) 25 (B) 30 *(C) 125 (D) 150 (E) 625

No Comments

Question 5, Test 2

Rater Mean 3.71
Perpendicular Distance -1.07

(See Table C5-question also identified by panel as favoring recent group).

Question 11, Test 2

Rater Mean 4.43
Perpendicular Distance -0.91

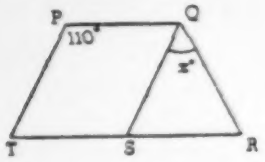
(See Table C5-question also identified by panel as favoring recent group).

Table C8

The Five Mathematical Questions with the Greatest (Positive) Perpendicular Distance (These questions were relatively more difficult for the more recent group).

Question 9, Test 1

Rater Mean 2.86
Perpendicular Distance 0.76



9. In the figure above, PQST is a parallelogram; QR = QS and TR is a straight line. If $\angle P = 110^{\circ}$, then x = (?)

*(A) 40 (B) 50 (C) 70 (D) 110

(E) It cannot be determined from the information given.

No Comments

Question 2, Test 2

Rater Mean 2.29
Perpendicular Distance 0.89

(See Table C6 question also identified by panel as being more difficult for recent group)

Question 8, Test 2

Rater Mean 3.00
Perpendicular Distance 0.83

8. A farmer has completed fencing all but one side of a square field. If he has already used 3n feet of fencing, how many feet will he need for the last side?

(A) $\frac{3n}{4}$ ft. *(B) n ft. (C) $\frac{4n}{3}$ ft.

(D) 3n ft. (E) 4n ft.

Comments on Question 8:

1. Use metric units.

Question 22, Test 2

Rater Mean 3.14
Perpendicular Distance 0.91

22. If $\frac{x}{2}$ is the perimeter of a square, then the area of the square is

*(A) $\frac{x^2}{64}$ (B) $\frac{x^2}{36}$ (C) $\frac{x^2}{16}$ (D) $\frac{x^2}{4}$ (E) $4x^2$

No Comments

Question 1, Test 3

Rater Mean 3.00
Perpendicular Distance 0.69

1. At a certain game, there were 48,000 spectators. If $\frac{1}{3}$ of the spectators bought programs and an average of 200 programs was sold by each program vendor, how many vendors were hired for the game?

(A) 32
(B) 40
*(C) 80
(D) 160
(E) 240

No Comments